AB-INITIO MATERIALS DISCOVERY AND CHARACTERIZATION THROUGH ENERGY LANDSCAPE EXPLORATION WITH AN EVOLUTIONARY ALGORITHM

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

William Wakefield Tipton Jr

May 2014

AB-INITIO MATERIALS DISCOVERY AND CHARACTERIZATION THROUGH ENERGY LANDSCAPE EXPLORATION WITH AN EVOLUTIONARY ALGORITHM

William Wakefield Tipton Jr, Ph.D.

Cornell University 2014

We present an evolutionary algorithm which predicts stable atomic structures and phase diagrams by searching the energy landscape of empirical and *ab-initio* Hamiltonians. Composition and geometrical degrees of freedom may be varied simultaneously. We show that this method utilizes information from favorable local structure at one composition to predict that at others, achieving far greater efficiency of phase diagram prediction than a method which relies on sampling compositions individually. We detail this and a number of other efficiency-improving techniques implemented in the Genetic Algorithm for Structure Prediction (GASP) code that is now publicly available.

Applications are presented in three categories. First, we predict phase diagrams of elemental barium and europium under pressure and show that our methodology compliments experimental studies of those systems. Second, we show that phase diagram prediction is a primary component of *ab initio* Li-ion battery electrode characterization. We present studies of the Li-Si and Li-Ge binary phase diagrams that allow us to determine the voltage characteristics of silicon and germanium battery anodes. We also predict the stability of previously-unreported binary structures in both of those materials systems. Third, we use the method to test empirical energy models. It is important that such models reproduce the energy landscape of the true system they are meant to represent.

The GASP code can verify this if it is so and find errenous structures to augment the fitting database if it is not. Our results suggest that genetic algorithm searches can be used to improve the methodology of empirical potential design.

This thesis takes advantage of the Cornell graduate school's "papers" option. That is, it is primarily composed from the author's first-author publications, in particular, Refs. [149, 148, 152, 151]. Additionally, one of the pleasures of computational materials science research is that it has a synergistic relationship with experiment and lends itself to many fruitful collaborations. These provide insights and richer publications than would be possible by either path alone. Thus this thesis also describes applications of our methodology to collaborative works described in Refs. [17, 145, 114]. In these cases, I focus on my own contributions in this thesis and refer the reader to the original publications for the full picture.

BIOGRAPHICAL SKETCH

Will was born in Nashville, TN on August 24, 1986. He attended high school at the North Carolina School of Science and Math where he met Yuki Jung and graduated in 2004. He then flew to Claremont, CA. to attend Harvey Mudd College. He intended to study computer science, but became interested in real science somewhere along the way and ended up as both a physics and math/CS joint major.

Will discovered computational materials science via a set of online video lectures by Gerbrand Ceder at MIT and spent the summer after his junior year working in Professor Ceder's lab. He found the work provided a great way to meld his love of practical scientific work with his actual skills which were computational. After graduating college, he left California for upstate NY to continue this path. He joined Richard Hennig's research group, and this thesis describes much of his work there. After graduation, he will head back to CA to begin as a software engineer at Google.

To my parents.

ACKNOWLEDGEMENTS

First, I'd like to thank Kent and Kiran, Houlong, Arunima, and CJ for being great labmates and making my time here fun as well as educational. Good luck to the younger generation of Hennig researchers, Ben and everyone else, who I know will go on to do great things. I've been lucky to work with a number of awesome undergrads including Stewart, Josh, Dan, and Michael. I'd like to thank Professors Shoemaker and van Dover for being on my committee and offering helpful feedback and suggestions. Finally, thanks most of all to Professor Richard Hennig for being an extremely supportive mentor and helping me grow as a scientific researcher.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

ENERGY LANDSCAPES AND STOCHASTIC SEARCH

Much of the content of this chapter was previously published in Ref. [149].

1.1 Introduction

A primary goal of computational and theoretical materials engineering is identification of materials with desirable properties. Often, we have an application in mind and can describe the properties of a material which may be successfully applied to our problem, e.g. it is light-weight, cheap, strong, insulating, or has specific band-gap or diffusion coefficients. It is then up to a materials engineer to find such a material.

To this end, we often approach the inverse problem. That is, instead of starting with a list of properties and working directly to a material solution, we start with a particular material and try to determine its properties. This is easier. Of course, once the properties of a long list of materials are known, it is likely we will be able to select from the list materials which satisfy the constraints of a given application.[71, 72] Today, parts of this problem are very routine calculations. *Once we know the atomic structure of a material,* methods such as Density Functional Theory implemented in a number of mature software packages allow us to predict a material's electronic structure, elastic constants, etc.[138] However, the question of how to find a material's atomic structure is an open one, and thus the need for the present work. Indeed, the question itself needs to be more precisely specified, since a material's structure may depend on growth conditions, its processing history, etc.

We know that at thermodynamic equilibrium, a material will take on the structure with the lowest free energy given by

$$G = U - TS + PV$$

where U, T, S, P, and V are the internal energy, temperature, entropy, pressure, and volume of the system, respectively. Since we compare trial solutions to each other in order to find the lowest energy structure, we are not as interested in the absolute free energy of any particular structure but in *differences* of free energies. The PV term is easy to find and include in a calculation and has significant effect on the results primarily when one is studying systems under high pressure. In practice, the internal energy U will account for most of the energy difference between phases (as well as most of the algorithm's run time). It may be calculated by way of energy models such as empirical potentials (using, e.g., GULP [56]) or density functional theory (using, e.g., PWSCF [12]).

We can break the entropy term S into three contributions: electronic, configurational and vibrational entropy [112].

$$S = S_{\text{el}} + S_{\text{conf}} + S_{\text{vib}}$$

The electronic term S_{el} is relatively easy to calculate but typically of negligible magnitude. The configurational and vibrational contributions can be significant but are difficult to compute as they require extensive sampling of the potential energy surface. For these reasons, the entropic terms in the free energy are often neglected. This is often safe to do, since we are primarily interested in differences in energies rather than the absolute quantities, and the entropic contributions cancel to some extent between different phases. However, there is the possibility of entropic stabilization in which structures which are not even mechanically stable at zero temperature can be stabilized entropically.[136, 141].

Nonetheless, due to the computational cost of directly estimating the free energy, stochastic-search algorithms are generally not applied to the high-temperature problem (Monte Carlo or Molecular Dynamics techniques may be useful here). We neglect the entropy for simplicity, essentially confining ourselves to the zero temperature regime. In this case, a material's free energy is simply its enthalpy $H = U + PV$, and a material's thermodynamically stable crystal structure is that arrangement of atoms which has the lowest enthalpy. Hence, to find the physically-realized crystal structure of a material with a particular composition, we must search for the atomic configuration with the lowest enthalpy.

In this light, we are viewing atomic structure prediction as an optimization problem. That is, if we view the energy of a system as a function of various parameters describing it (atomic positions, etc.), then predicting the stable structure is equivalent to finding the values of the parameters that minimize the energy function. Unfortunately, the energy functions of real systems are not simple objects, and we cannot write analytic expressions for them. These functions themselves are expensive to compute, and derivatives thereof are as well. Additionally, they are known to have many local minimums. These two properties of the objective function force us to exclude most traditional optimization methods. Deterministic global search strategies such as branch-and-bound may be considered, but they remain exceedingly computationally expensive even with the simplest of energy models.[44]

Any optimization problem for which we can describe and evaluate solutions is amenable to one of the most simple optimization schemes: guess and check. We guess a variety of possible solutions, evaluate the quality of each of them,

and choose the best. There are many ways one can imagine to guess solutions including researcher intuition, but one of the most simple choices is the topic of this chapter: random search. Additionally, we will see that local minimization routines improve random searches beyond the trivial guess and check methodology. We begin the remainder of this chapter by briefly discussing the history of random search methods and their application to crystal structures. We present an overview of the theory motivating the design of most methods. Finally, we survey the method's application to various systems of interest in the literature.

1.2 History and Overview

There are many optimization problems of great practical importance, and random search algorithms have long been applied to their solution. The method was probably [140] first suggested by Anderson [7] in the context of operations research and further investigated shortly thereafter by Rastrigin [129] and Karnopp [77]. More recent texts by Spall [142] and Zhigljavsky [181] provide a comprehensive discussions of stochastic search and optimization methods. Before looking at the details of any particular implementation of a random search algorithm for atomic structure prediction, we make some general comments about the method.

Given some details about our system of interest, say the stoichiometry of a solid crystal or the sequence of a protein, the random search program is to repeatedly generate some random arrangement of the system's atomic or molecular components and, subject to local minimization, compute the energy of that arrangement. This is repeated until a sufficient solution is found. We can dia-

gram this simple algorithm as follows.

1. Randomly generate structure
2. Apply local optimization routine to minimize structure's energy
3. Repeat until convergence

Randomly guessing solutions may notbe the most efficient way to solve a problem, since it does not leverage any *a priori* knowledge of the problem, nor anything we learn about it over the course of the search. However, the simplicity of this approach which is the root of its weaknesses also leads to several advantages. First, the method is relatively quick and easy to implement. As computer time becomes less and less expensive relative to programmer time, navigating this trade-off becomes more important. Second, the method need only use the energy of particular structures. Random search methods commonly use low-order derivatives, i.e. forces and stresses, for local optimization, but they are not required, and so the method is compatible with energy routines which either can not produce such extra information or which are prohibitively slow in doing so.

Next, although determining convergence to the absolute ground state is challenging, we find in practice that the method is often very quick to find good solutions, energetically low-lying configurations which bear many similarities to the thermodynamic minimum and which may occur as metastable phases in the material. As we will see in specific cases below, with the significant reduction in problem complexity afforded us by local optimization, we can also often be confident that the algorithm has found the global minimum.

Finally, the method is easily amenable to statistical analysis for describing its convergence properties. Anderssen and Bloomfield show that random searching is more effective than searching on a uniform grid in phase spaces of dimensionality greater than six [8]. Spall shows how to estimate the number of trials required to assure a certain probability of sampling the correct solution. He assumes that a finite volume of the solution phase space corresponds to the optimal solution (such as the basin of attraction of the local minimizer in our problem) and derives an expression for the number of samples required to guarantee a certain probability of finding it [142]. See [140] for discussion of the method's convergence rate and [142] for additional mathematical details. More on the method's convergence rate and mathematical details can be found in Refs. [140] and [142], respectively.

The most computationally expensive step in this process is the energy calculation. Although the details of the system, algorithm, and energy method are relevant to the method's success, accurate energy calculations typically take several orders of magnitude more computer time than any other step in the algorithm. For this reason, run-time of the algorithm may be described by the number of total energy calculations which must be performed to achieve convergence, and it is important to minimize this number.

1.3 Methods

It is necessary to specify the form of trial solutions. This choice has wide consequences for the success of the algorithm. Most work on structure prediction by random search has focused on crystalline solids. We will concentrate on that

problem. Similar considerations will apply when parameterizing other types of systems such as molecules or nanoclusters.

By assuming periodic boundary conditions, an infinite crystal may be specified by a Bravais lattice and a basis of atoms. The lattice may be specified by six lattice parameters (such as three angles and three lengths), and the basis by N atomic coordinates (i.e. $3N - 3$ numbers for 3D crystals, taking into account translational invariance). Thus, the solution phase space for this problem is of $3N + 3$ dimensions. The number of atoms in the basis, N, is a parameter which itself may need to be determined by the search algorithm.

The solution space we have described is high dimensional and infinite in extent in most of the dimensions. However, many points in the phase space either represent clearly unphysical structure or are redundant, describing crystals which are also described by other points in the space or represent crystals which may be immediately excluded as unphysical. By designing some simple criteria that eliminate the obviously-nonphysical structures and confining our search to a single representation of each crystal, one can avoid considering large portions of the total solution space and simplify the search problem.

Unphysical solutions include those which contain atoms spread very far apart. By applying constraints on the parameters of our trial solutions: minimum and maximum lattice parameters, we confine our search to a bounded space. We may also constrain the range of allowable nearest-neighbor distances or crystal densities to further narrow our search to physically-realizable structures. Incidentally, this often also helps to ensure the stability of energy codes.

Now, notice that even the bounded solution space is highly redundant. A

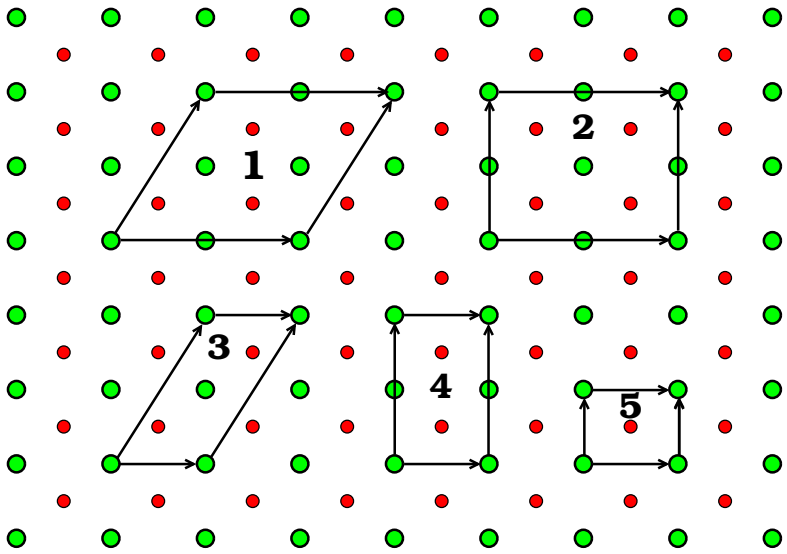


Figure 1.1: Five alternate representations of a single physical crystal are shown. Cells 2 and 4 are Niggli reduced versions of cells 1 and 3, respectively. They are also 2x2 and 1x2 supercells, respectively, of the primitive cell 5.

single physical crystal can be represented by many different unit cell choices, by shifting all the atoms by some constant amount, or by swapping the coordinates of two identical atoms in the basis. Figure 1.1 illustrates this representation problem for the case of a crystal structure. For clusters or molecules, rotations are also redundant degrees of freedom. It is important that search algorithms attempt to represent each structure in a single, standardized way. In this way, the algorithm may avoid redundant calculations and, again, significantly reduce the effective size of the solution space. We may also constrain trial solutions based on experimentally-known data such as the space group or structural motifs (e.g., the H_2O units in ice).

To this end the algorithm generally enforces that a particular species is located at the origin in crystal coordinates and that the lattice itself is chosen in a standardized form. One such form is described by Pauschenwein.[115]

He presents the construction of "a general parameterization for all three-dimensional crystal lattices... which guarantees that the three primitive vectors constructed by the parametrization are the three shortest possible, linearly independent lattice vectors existing in the whole lattice" [115]. This Minimum Distance Parameterization removes almost all of the redundancy in the phase space.

The most important technique for reducing the complexity of the search problem is local relaxation of trial structures. While the traditional optimization algorithms built into most energy codes can not automatically find the *global* energy minimum, they can efficiently *relax* a given structure to a nearby *local* minimum. The particular methods used include conjugate gradient optimization, several quasi-Newton methods, and damped molecular dynamics [127] [47]. Essentially, the energy determines the forces acting on each atom and moves it "downhill" until the system reaches a minimum of the energy surface. Notice that this will not usually overcome any energy barriers between the trial solution and the true ground state. This is illustrated in Fig. 1.2 using a one-dimensional function. Note that the potential energy surfaces of real systems lie in spaces of high dimensionality which leads to complications not apparent in the 1D case.

Performing the relaxation algorithm on each trial structure before calculating its energy significantly reduces the complexity of the structure search problem. Instead of randomly guessing the ground state solution itself, the method must only guess a solution "nearby" the ground state. We may think of this as partitioning the solution phase space into regions of attraction of the local optimization method surrounding each local minimum. This is illustrated in Fig.

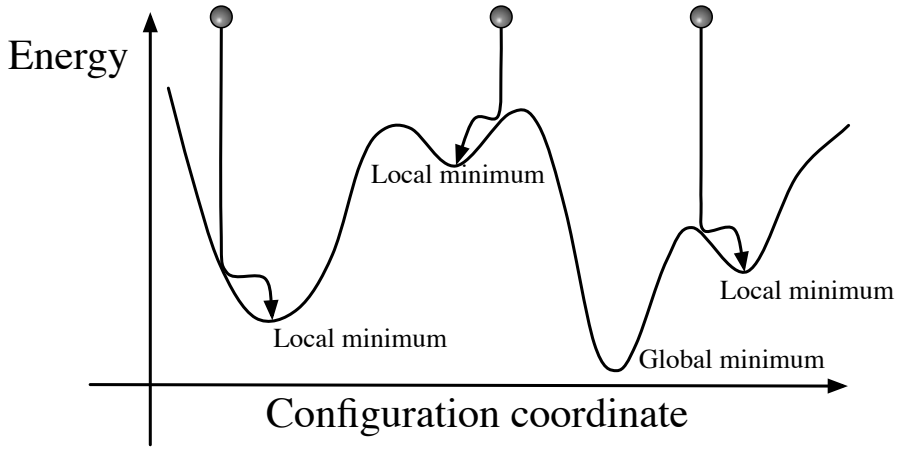


Figure 1.2: The random search process using local minimization. The algorithm randomly generates a variety of trial solutions which are relaxed to a nearby local minimum using a traditional optimization routine.

1.3. To completely search the solution space, we must no longer sample every point in the space, but merely one point in each region.[160]

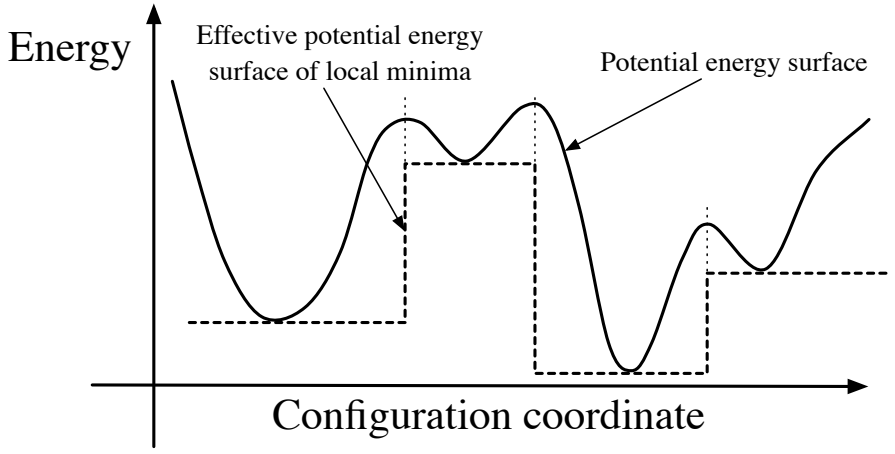


Figure 1.3: Effective potential energy surface. The use of a local minimizer simplifies the search problem by transforming the continuous solution space with infinitely-many trial solutions into a discrete space with finitely-many if we constrain our search space to a finite volume using physical constraints.

With this technique, we may begin to speak with some confidence of exhaus-

tively sampling a space. A properly constrained solution space may have few enough local minima that a search algorithm may sample all of them in a reasonable amount of time. Several authors suggest that their random searches may have indeed been nearly or fully exhaustive since they found many structures several times each.[118, 50].

However these works considered relatively small unit cells. As per our discussion above, the size of the solution space and the complexity of our problem may be quantified by the number of local minima in the constrained space. The dimension of the search space grows linearly with the N. Therefore, it is believed that the number of local minima in the solution space and, thus, the number of trial solutions we need to adequately sample it, grows exponentially with N.[16, 144].

Finally, a search method must specify criteria for algorithm convergence. By the nature of the method, it can never find the ground state solution with absolute certainty. This limitation is, of course, common to any other search method, such as genetic algorithms. In practice, several naive convergence criteria (or "stopping criteria") work well. A search is usually considered converged when no improvement in the best trial solution has been made over several iterations and the current best solution has been found several times.[35] Venkatesh et. al. have developed a more sophisticated statistical method to this end.[158] The method applies Bayesian analysis to the set of local minima found over the course of the random search to approximate the distribution of the number of local minima. From this, they derive the convergence criterion which tries to navigate the computer time/solution confidence trade-off.

1.4 Applications and Results

Due to their periodic structure, simple crystalline solids generally have many fewer degrees of freedom than do, e.g. proteins. Thus, work making use of random structure searches has focused on these simple systems where the computer time/programmer time trade-off is appropriate. In all cases, the energy and local minimization routines used are standard, so a work's random structure generation method and convergence criteria are its salient features. Some of the earliest work on structure prediction by random search gave a proof-of-concept on Lennard-Jones systems.[135] The following works describe real materials.

The Pickard and Needs group from Cambridge has been very prolific in their use of the random search strategy in recent years, having investigated a number of interesting molecular crystal and semiconductor systems described by Density Functional Theory. In 2006 they applied their method to silane (SiH_4) at high pressures in order to find phases which may superconduct.[118] They argue that the method is particularly suitable for systems under high pressure since such phases often have simple structures in the sense that they frequently have small unit cells. They perform their search separately for cells of different numbers of atoms, N. Once the number and type of atoms in the cell is chosen, they stochastically generate a lattice by selecting three lattice lengths distributed uniformly on $[0.5, 1.5]$ and three lattice angles distributed uniformly on $[40°, 140°]$. The volume of the entire cell is then scaled to between 0.5 and 1.5 of some given volume, and Si and H atoms are given random coordinates uniformly distributed in the cell. They found a metallic phase which should be accessible experimentally.

In the same year, the group studied high pressure structures of CaC_6, a material whose superconducting properties vary with pressure.[35] Initial structures for the search were specified to contain 7 atoms and the lattice randomly generated such that the density was within a factor of two of the known low-pressure phase. The authors suggest that the search was likely nearly exhaustive (over structures with $N = 7$) since they found many phases several times. They used intuition gained from the 7-atom case to construct and test structures of larger N. The study yielded several structures which were favored at different pressures and which were later experimentally confirmed.[43]

The same authors have modified the random search technique to find the structure of hydrogen defects in Silicon. This was done by enforcing additional constraints on the trials solutions, i.e. each trial solution contained a defect. They found novel structures for defect clusters of various sizes.[102] Similar studies of high-pressure phases of solid hydrogen[120], nitrogen[124], lithium[123], H_2O [121], aluminum hydride[119] and iron[125] yielded novel results. A review by Pickard and Needs of their own work is given in [122] and provides additional insight into their approach.

Feng et. al. performed an extensive structural search of Li-Be compounds using both random search and additional guess and check based on chemical intuition [50]. The random search was instrumental in identifying stable high-pressure phases where the researchers' intuition was less successful. They found four stoichiometric Li_xBe_{1-x} compounds stable over a range of pressures, several of which display quite unusual and unexpected electronic properties. In Fig. 1.4, we show the structures and the pressure ranges over which they are stable.

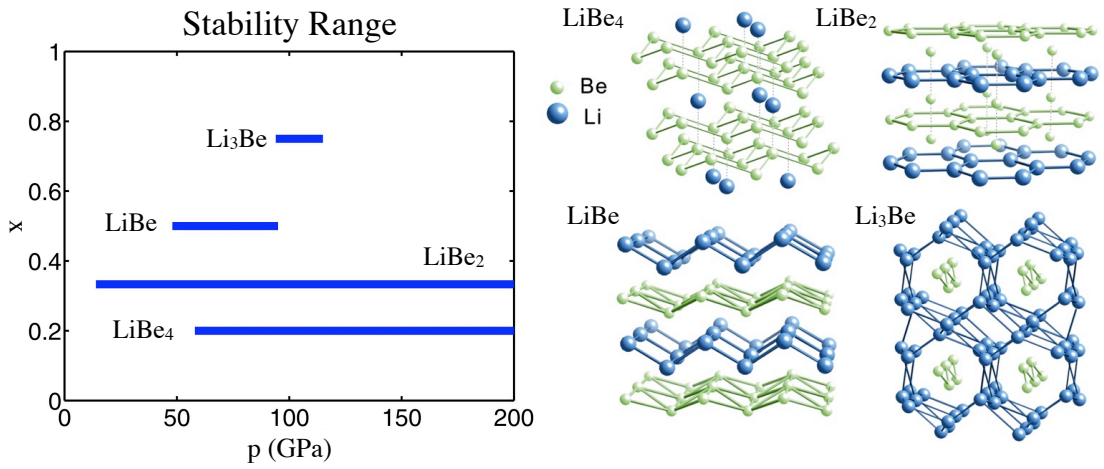


Figure 1.4: Li-Be compounds found at high pressure using a random search method by Feng et. al. [50].

Successful random searches have been performed on metal and metal-alloy nanoclusters by Johnston et. al. In Ref. [92] they study aluminum clusters described by the Murrell-Mottram potential. Trial clusters are randomly generated subject to constraints on minimum and maximum nearest-neighbor distance and the constraint that all atoms lie within a sphere of radius proportional to the number of atoms $N^{\frac{1}{3}}$. The clusters are then relaxed using a quasi-Newton method. In this way, 1000 trial solutions were prepared and tested for clusters of each size $N = 2 \ldots 20$. In most cases, the ground state is found within the first 100 trials and on the order of 10 times in total. Additionally, they describe evidence for exhaustive sampling in some of these clusters. However, they also find that the method becomes less effective with increasing system size.[92]

In 2003, Bailey, Johnston, et. al. studied Ni-Al alloy clusters using the Gupta empirical potential with both random search and genetic algorithm techniques [11]. They generate structures in the same way as in the work described above.

Although both methods are successful for small cluster sizes, they find their genetic algorithm to be more efficient than random search, especially in alloy systems which have more degrees of freedom.

Larger systems such as proteins have relatively complicated phase spaces, and prediction of their structure often benefits from specialized techniques.[161, 134] However, Wang et. al. found that the efficacy of their genetic algorithm approach was improved by combining it with a random search method. The hybrid method is implemented by replacing some proportion of the worst solutions in each generation by randomly-generated ones. This increases genetic diversity in the population and avoids premature convergence to non-global minima more effectively than evolutionary mutation operators.[166] The introduction of random structures has played a role in maintaining genetic diversity other evolutionary algorithms as well.[169]

Several tests of random searching applied to small organic molecules have been made in the context of a comparison to other methods.[45, 46]. The comparisons found random searching (as well as most other methods tested) rather ineffective in predicting the structure of several molecular systems. However, these tests fail to take into account one of the largest benefits of the random search method: simplicity and short amount of researcher time to solution. More importantly, the optimization methods tested made use of a variety of energy codes but were evaluated based on whether or not they found the experimentally-known solution. This is problematic since many of the objective functions being optimized may not have even had a global minimum at the physically-correct structure. Thus, whether or not a particular optimization method found the experimental structure is a poor indicator of its success.

These are interesting works that look at overall strategy but do not separate energy model from optimization scheme in their method evaluation.

1.5 Summary

We have reviewed the theory of random search techniques and their application to materials' atomic structure prediction in the literature. The method has been successfully applied to many interesting and technologically-important systems. For binaries and other systems with relatively small solution spaces, the method is reliable and has the significant advantage of requiring very little programmer time to obtain a solution. More complicated systems may benefit from the use of methods such as genetic algorithms which incorporate information learned about the system over the course of the algorithm to make better guesses.

CHAPTER 2

A GENETIC ALGORITHM FOR STRUCTURE PREDICTION

Much of the content of this chapter was previously published in Ref. [151].

2.1 Introduction

Knowledge of a material's atomic structure is a prerequisite to many computational materials studies. Once the structure is known, a variety of theoretical methods allow us to predict its mechanical properties, electronic properties, etc. The computational prediction of materials' structures is a long-standing problem[99]. Recent advances in *ab initio* structure prediction have enabled the discovery of novel materials and the crystal structure solution of known complex materials.[118, 50, 17, 182, 52]

Structures which are thermodynamically stable are those which lie at the minimum of a free energy function. Thus, structure prediction can be framed as an optimization problem and solved by searching for structures with the lowest free energies. No direct solution of this minimization problem is possible, so it has been addressed using heuristic optimization methods. Pickard *et al.* provide a good discussion of the issues surrounding this optimization problem in the context of a random search method.[126] Genetic algorithms have received much attention and have had many successes.[3, 22, 37, 73, 57, 106, 156, 169, 93, 49, 154, 172, 173] Methods such as particle swarm optimization,[167] minima hopping,[5] data mining methods,[25, 51] and others have also produced promising results.[149]

The success of these methods clearly relies on the ability to perform accurate free energy calculations. However, these calculations are logically separable from the optimization problem; the optimization is successful if it is able to identify the global minimum of any energy function it is given, regardless of whether that energy function faithfully reproduces the physics of any particular system. In practice, the free energy is often approximated by an internal energy or enthalpy obtained within the density functional theory (DFT) formalism, neglecting entropic contributions, although there is no reason in principle why additional contributions to the free energy could not be included during the search procedure. In any case, the energy calculation is generally treated as a black-box by the optimization algorithm. In this chapter, we focus on the optimization problem, and we will generally use the term energy to refer to whichever thermodynamic potential we happen to be minimizing.

The general approach of heuristic structure optimization methods may broadly be described as guess-and-check. Candidate solutions are generated and then evaluated using an energy model. Their values are compared to those of previously encountered candidates. In the simplest approach, described in the previous chapter, a search algorithm might generate candidate solutions more-or-less randomly. A more sophisticated search method must leverage *a priori* information about the problem and/or "learn" about the problem over the course of the algorithm. The genetic algorithm described in this chapter will do both of these things. We may visualize this by imagining that the method begins with a broad sampling of the solution phase space and then makes use of the information it gains to focus in on the more promising regions. The objective function, known as the energy landscape, is defined over our solution phase space and is complex and high-dimensional.[126, 157]

For many applications, the result desired is a prediction of the most stable structure at a particular composition, and so searches are generally performed by considering only candidate solutions at the composition of interest. In this case, the structure with the lowest energy per atom encountered is predicted to be stable. This standard methodology is an over-simplification of the structure prediction problem in at least one respect. The condition that a structure is at a minimum of the energy functional for a given composition is not a sufficient condition for it being thermodynamically stable.[50, 175] Indeed, a naïve search algorithm will predict a lowest-energy candidate at any given composition, but for most compositions, no single stable crystal structure exists in nature. Instead, the true ground state could be a mixture of multiple phases.

To see why, consider the zero temperature and pressure Zr-Cu binary phase diagram in Fig. 2.1. Each point in the figure represents a structure and is plotted according to its composition on the abscissa and its energy per atom on the ordinate. The line connecting the lowest-energy structures is known as the convex hull and all thermodynamically-stable structures lie on this line.[52] Consider, for example, the composition $X_{\mathrm{Zr}} = \frac{5}{6}$. At this composition, the most stable material which can be formed is a material which is actually a 50-50 mixture of the crystal structures which form at compositions $X_{\mathrm{Zr}} = \frac{2}{3}$ and $X_{\mathrm{Zr}} = 1$, that is, of those which form the endpoints of the line segment of the convex hull found at $X_{\mathrm{Zr}} = \frac{5}{6}$. A heuristic search algorithm constrained to a small number of atoms and the composition $X_{\mathrm{Zr}} = \frac{5}{6}$ would identify a lowest-energy candidate. If we plotted that structure in Fig. 2.1, it would lie above the convex hull and the structure would be unstable with respect to decomposition into a mixed phase of compounds with different compositions. The vertical distance of the energy of the structure above the convex hull, ΔE_0, corresponds to its formation energy

with respect to the true ground states.

Therefore, even if one is only interested in finding the material which forms at a specific composition, *ab-initio* structure prediction in the case of a multi-component system require a search over the materials system's entire composition space. This amounts to solving for the complete constant temperature and pressure phase diagram, a capability which is desirable for many other reasons as well.

In this chapter, we describe our implementation of a genetic algorithm software package, the Genetic Algorithm for Structure and Phase Prediction (GASP), to solve the structure and phase diagram prediction problems. The phase-diagram searching extension is based on the method of Trimarchi *et al.*[154] The GASP software package is freely available and is interfaced to a number of energy and local optimization codes.[150]

To test the genetic algorithm, we will apply it to the Zr-Al-Cu system using an embedded atom method (EAM) potential designed by Cheng *et al.*[30] This energy model has been leveraged by many other authors in order to study this technologically-important materials system.[178, 164, 174, 165, 179, 55, 163, 41, 91, 24, 28, 6, 39, 177, 40, 54, 29, 176, 9, 170, 67, 83] We usually perform energy calculations at the DFT level to obtain production-quality predictions, and we will describe such search in future chapters. Here, however, we search over the energy landscape defined by the Zr-Al-Cu EAM potential. The search method is stochastic, so drawing statistically significant conclusions about its success involves many repeated runs, and *ab-initio* energy calculations are prohibitively expensive for this purpose.

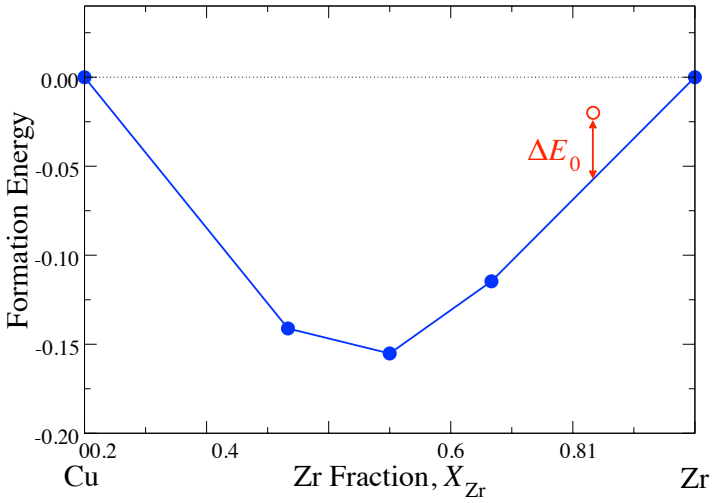


Figure 2.1: Illustration of the formation energies for the Zr-Cu phase diagram using an embedded-atom potential. The ground state structures correspond to the points forming the convex hull. At compositions between the ground state structures, the thermodynamic ground state is a mixture of the two neighboring phases. For example, there is no stable single phase with composition $X_{Zr} = \frac{5}{6}$. Instead, the ground state is a 50-50 mixture of the crystal structures with compositions $X_{Zr} = \frac{2}{3}$ and $X_{Zr} = 1$.

2.2 The energy landscape

The hypersurface we want to minimize has often been referred to as the energy landscape. This is meant to evoke a geographical landscape of hills and valleys. In this picture, our goal is to find the deepest valley. However, unlike in the geographical problem, our objective function is very high dimensional. Each structure, a candidate solution to the energy minimization problem, may be represented by $3N + 3$ coordinates: 6 lattice parameters and $3N - 3$ atomic coordinates. (The number of atoms N of the system is also a free parameter which must be determined.) It has been shown that the number of local minimums increases exponentially with N. 80,91

That said, as a consequence of the No Free Lunch theorem, we must understand and take advantage of certain properties of the energy landscape in order to perform better than a random search. We summarize here the physically-motivated properties we expect and which the GA will attempt to leverage:

- Depending on the system, interatomic distances of smaller than about 1 or larger than about 4 are usually unphysical. Atoms closer than that lower bound will strongly repel one another, and nearest neighbor distances above that upper bound are generally not ground states either. This observation allows us to exclude large portions of the original solution space.
- Atoms near one another tend to interact much more strongly than those far apart. Thus, there is a partial spatial separability to our problem – a system composed of spatially-coherent regions, each of which is favorably-configured, is likely to be a low energy system overall.
- Any local minimization routine, such as steepest descent, implicitly partitions the energy landscape into basins of attraction of the local minimizer. Since we can assume we are able to relatively efficiently perform local minimization, we need only find a structure in the basin of attraction of the global minimum. Furthermore, deeper basins tend to occupy larger volumes in our solution phase space. In particular, Massen *et al.* find an exponential relationship between the depth and volume of a basin.[101] This greatly simplifies the search problem.
- Low energy minima usually have high symmetry and repeated structural motifs.[126]
- Low lying minima are usually located near each other in the energy landscape. In particular, the global minimum is often surrounded by other low

lying solutions. [108]

Thus, the energy landscape has an overall structure which can be exploited by a search algorithm.

2.3 Genetic algorithm

Our genetic algorithm (GA) implementation makes use of many features which have already been described in recent literature. Thus, we begin by briefly outlining our approach and pointing out the origins of various ideas and components. We then turn to a more detailed discussion of several aspects of our algorithm which are unique or which we believe are important to its success and have not been given sufficient treatment.

The evolutionary approach to crystal structure prediction is modeled after the biological process of evolution. Each structure, a candidate solution to the energy minimization problem, is analogous to an organism. Its representation in $3N + 3$ coordinates will be its "genotype". The number of atoms N of the system is, for computational feasibility, generally limited to some value less than 100. For crystal structure searches we assume periodic boundary conditions, i.e. enforce the crystallinity on our solutions, allowing us to study bulk material. In nature, the fitness of an organism is based on how well its phenotype is suited to its environment and, in particular, how successful it is in reproducing. We assign fitnesses to crystal structures based on their energies, and we allow them to "reproduce" probabilistically based on those fitnesses.

Figure 2.2 illustrates the algorithm. Candidate structures, or organisms, are

organized into groups called generations. The algorithm proceeds by creating successive generations. The initial population is composed primarily of structures whose lattice parameters, number and type of atoms, and atom positions chosen randomly from uniform distributions bounded by the hard constraints on the candidate solutions' geometry described in Section 2.3.4. Specific structures may also be included in the initial generation if desired.

The operations by which an offspring generation is made from parents are called variations and, inspired by biology, consist of mutation and mating operations, which are described in detail in Secs. 2.3.2 and 2.3.3. In order for successive generations of structures not to lose information about the best solutions encountered, we implement a promotion operation which copies some of the best organisms. In the case of single composition searches, this number is generally a free parameter. In the case of the phase diagram search, all structures on the convex hull are generally promoted from one generation directly to the next.

When searching for the lowest-energy structure at a single, fixed composition, an organism's fitness is inversely related to its total energy per atom. Evolutionary pressure analogous to that which forces species to adapt to their environments will lead to lower energy structures as long as two primary criteria are satisfied. First of all, better parents must be more likely to reproduce. The method by which we select structures to act as parents is extremely important, since this is the primary way we apply evolutionary pressure on the population to improve. We discuss this further in Section 2.3.1.

Secondly, the reproduction operators must be capable of passing down from parents to children whatever traits of candidate solutions are important for the

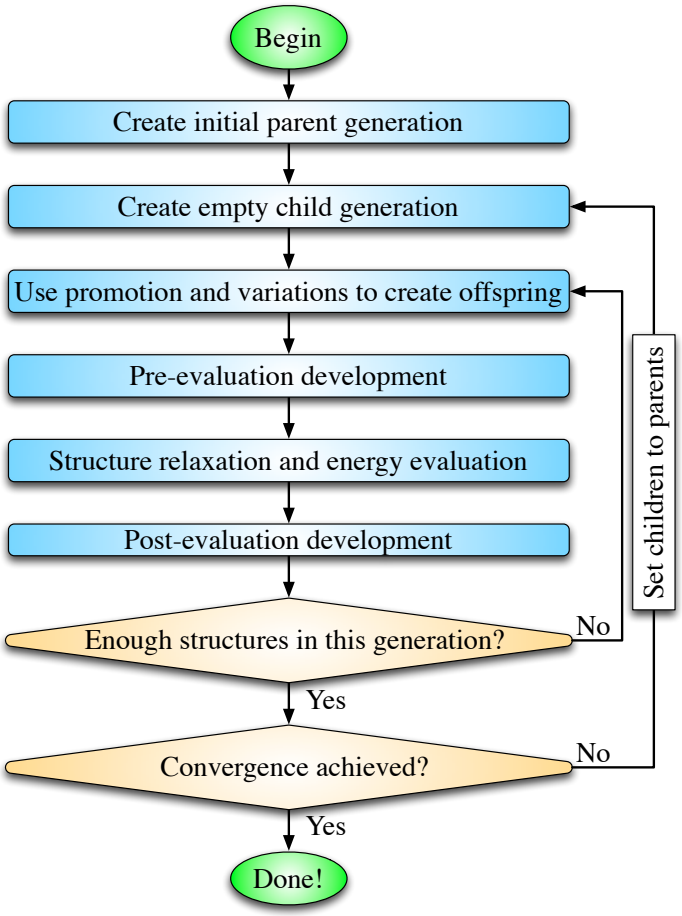


Figure 2.2: Outline of the basic genetic algorithm for structure prediction as implemented in the GASP code, neglecting the parallelization.

energy minimization problem. In the language of biology, the traits which control the organism's fitness should be highly heritable for the evolutionary process to be effective. That is, children must "look like" their parents. Both of these criteria must be satisfied in order for the evolutionary process to effectively act as an optimization procedure. We will discuss how our method satisfies this criterion in Section 2.3.3.

While these two conditions for a successful genetic algorithm may seem obvious, they are not always appreciated. For example, traditional genetic algo-

rithms often represent candidates solutions as binary strings and perform their mating operation by splicing these strings. Since binary strings do not provide a representation that focuses on energetically favorable structural motifs, it is unlikely that the traits which characterize a low energy structure are passed from parent to child. As a result, there is little reason to expect a GA using binary strings to perform better than a simple random search, and it could easily perform worse. More generally, it is important to notice that the genetic algorithm is really more of a general problem-solving strategy rather than a specific algorithm. Many design choices are possible, and details of the implementation significantly affect the success of the method.

2.3.1 Selection

Each time the algorithm generates a new organism using a variation, it selects one or more structures to act as parents based on their fitnesses. Preferential selection of more fit parents is the major evolutionary "force" that drives the population to improve. An organism's fitness is its value as evaluated by the objective function (e.g. its energy per atom), normalized in the context of its generation. In particular, an organism with an objective function of value v is assigned a fitness f given by

$$f = \frac{v - w}{b - w},$$

where w and b are the objective function values of the worst and best structures in the current generation, respectively. So, the best organism has a fitness of 1 and the worst has a fitness of 0, and we maximize fitness in order to minimize energy.

There are several commonly-used selection strategies. In truncation selection, some fraction (say, a third) of the best organisms in the generation are all equally likely to reproduce.[73] In roulette wheel selection, the probability that an organism is selected to act as a parent is directly proportional to its fitness.[3] In this way, it is possible for any organism to reproduce (except for the worst one which has fitness 0), but it is more likely for organisms of higher fitness to be selected. Finally, in tournament selection, a subset of organisms in a generation is chosen randomly, and then the best of those is selected to reproduce.

We have designed a more general method: organisms are selected based on a probability distribution over their fitnesses. For the probability distribution we chose a power law. Two parameters are specified to describe the distribution: the number of parents N_{parents} and an exponent P. The best N_{parents} organisms have a non-zero selection probability, and the exponent P specifies the power-law for the probability distribution. We denote the different selection strategies by an ordered pair of these two numbers $\left(N_{\text{parents}}, P\right)$.

To select a single parent, we first set the selection probabilities of all organisms that have a lower fitness than the $N_{\text{parents}} + 1$ best to zero. We then recalculate the fitnesses of the remaining organisms with respect to the remaining sub-generation. The selection probability of a remaining organism with renormalized fitness f_i is set to

$$p_i = \frac{(f_i)^P}{\sum_j (f_j)^P},$$

where the sum in the denominator is over the top N_{parents} organisms to ensure that the probability distribution is normalized.

Figure 2.3 shows several examples of selection probability distributions. First of all, truncation selection can be achieved by choosing a relatively small

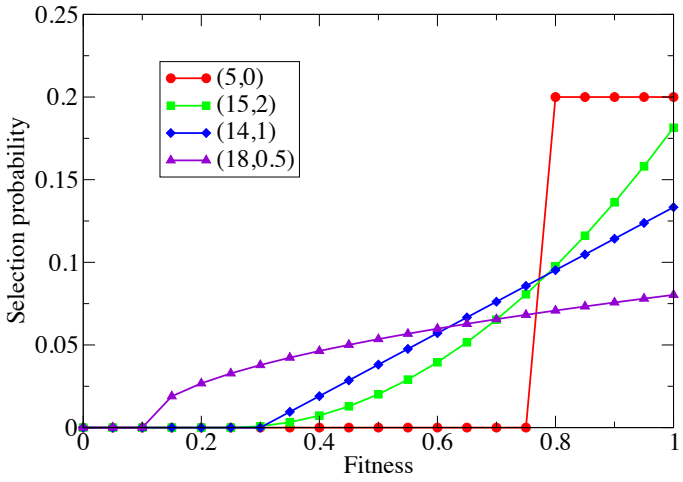


Figure 2.3: Example selection probability distributions of a generation with 21 members with evenly-spaced fitnesses. In order of aggressiveness, we have a $(5, 0)$ distribution in red, a $(15, 2)$ selection in green, a $(14, 1)$ selection in blue, and a $(18, 0.5)$ selection in violet.

N_{parents} and setting P to zero, and roulette wheel selection is equivalent to using a $P = 1$ and setting N_{parents} to the total generation size. Many other distributions are possible.

By varying the parameters describing the selection probability distribution, we can tune the amount of evolutionary pressure that we put on the population. For example, if we parameterize the selection algorithm so that it strongly favors the best few individuals in the population, then these few structures will often be chosen as parents. The area in solution phase space in the immediate vicinity of these will be searched thoroughly in a relatively small number of generations. The algorithm is thus likely to quickly converge to the lowest-lying structure nearby the best structures from the initial generations. However, there is significant danger that this convergence is premature, i.e. it is not to the global minimum. If, on the other hand, the selection routines are parameterized to favor only slightly the better structures over the worse ones, then the algo-

rithm will take longer to converge, but it will have searched the entire solution phase space more thoroughly when it does. Thus, our approach to parent selection enables more flexibility than traditional selection methods in navigating the important trade-off between computation time and confidence in the final solution.

2.3.2 Mutation

We consider four types of mutation operations: (i) perturbation of atomic coordinates, (ii) perturbation of lattice parameters, (iii) the random addition or removal of atoms, and (iv) the swapping of atomic locations. Depending on the specific application of the genetic algorithm, we may only use a subset of these operations and generally try to use mutation operations that are likely to introduce useful new information into the gene pool. For example, the random swapping of two atoms' positions is somewhat likely to lead to a good structure in the case of metallic alloys but not in binary ionic systems.[3]

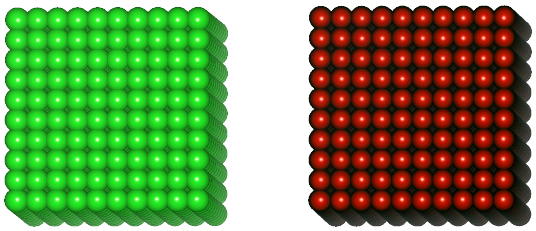
2.3.3 Mating

The energetically-important interactions in most materials are short-ranged. This suggests that there is some amount of spatial separability in the energy-minimization problem – low energy structures can often be thought of as composed of local structural motifs which are themselves energetically favorable. This observation motivates the most commonly used mating operation in which a coherent segment or real-space slice of one parent structure is extracted and

stacked on top of a similar segment from another parent. This operation maintains much of the local atomic structure from each parent, and thus the children tend to "look like" the parents in a respect pertinent to the energy minimization problem.[37]

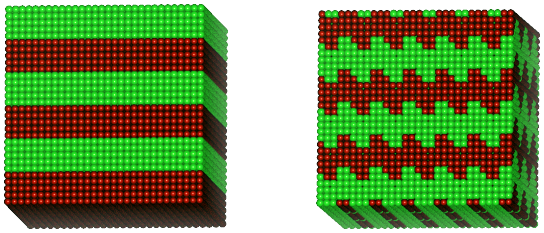
Figure 2.4(c) illustrates the mating operation applied to two artificial simple cubic structures. The mating variation first selects two parents. The lattice parameters of the offspring are the average of those of its parents. To decide which atoms from the parent cells are copied to the child, one of the three lattice vectors, A, is randomly selected, and a fractional coordinate s with $0 \leq s < 1$ is drawn from a uniform distribution. Then a slice thickness t with $0 \leq t < 1$ is drawn from a truncated Gaussian distribution. All atoms in one parent whose *fractional* coordinates along A are within $t/2$ of s are copied into the offspring structure. Atoms in the other parent whose coordinates along A are further away than $t/2$ from s are also copied into the offspring. Each site in the offspring has the same element type and fractional coordinates as an atom in the respective parent.

Several extensions to the mating operation have been proposed. The first, due to Oganov *et al.*, involves shifting all atoms in one parent structure by the same amount before mating.[57] These shifts may happen with different probabilities along the axis where the cut is made and along the others. This removes any bias caused by the implicit correlation between the coordinate s on the axis A in one parent with the coordinate s on the axis A in the other. In practice, this helps repeat good local structure to other parts of the cell. On the other hand, as we will discuss in Sec. 2.3.5, we do not necessarily want to encourage the replication of similar crystals with different representations in the population.



(a) Artificial cubic parent structure.

(b) Artificial cubic parent structure.



(c) Child created by the slicing mating operation using a horizontal cut.

(d) Child created by the slicing mating operation using a periodic cut.

Figure 2.4: The slicing variation. The top two structures are used as parents. The bottom two structures are the result of mating using different parameters. 3x3x3 supercells of the child structures are shown.

A second generalization is periodic slicing.[3] In this case, the fractional coordinate s, described above, becomes a cell-periodic function of the fractional coordinates along the two axes perpendicular to the A. We use a sine curve whose two amplitudes and wavelengths are pulled from uniform distributions. An example of the result of this sort of variation applied to two artificial structures is shown in Fig. 2.4(d).

2.3.4 Development

The development stage of the algorithm comes between an offspring's creation and it being added to the generation. It may be thought of as an organism's "growing up," and it is possible for the organism to fail this process and thus be discarded. In practice, this stage is responsible for performing the redundancy checks described in Sec. 2.3.6 and for the enforcement of the following hard constraints: minimum and maximum lattice parameters and number of atoms, minimum interatomic distance, and minimum number of species. These values may be chosen conservatively as to not bias the search but still avoid many unphysical or multiply-represented structures. Sane values of these parameters are also necessary to ensure the stability of many energy codes.

Additionally, we maintain an estimate of the optimal atomic density of structures independent of the population.[57] The density estimate is optimized by starting from an initial guess and then updating its value each generation. Our particular update scheme requires two parameters, a weight w and a number n. Each generation, we calculate the average density, ρ_g, of the top n organisms in the generation. The next optimal density estimate is then given by

$$\rho_{i+1} = w\rho_g + (1 - w)\rho_i,$$

where ρ_i is the current optimal density.

The volume of new structures is scaled to this density before relaxation. The reason is a practical one. Many local minimization algorithms are slow or unstable if the initial configuration is far from a minimum. This scaling is an easy first pass at moving a solution towards a local minimum.

2.3.5 Structure representation

A periodic structure with an N atom basis can be described by $3N + 3$ continuous variables (and N itself must usually be determined). However, these are not truly independent variables. A supercell, an alternate choice of lattice vectors, or any translation of the atomic coordinates may produce an equivalent representation of a particular crystal structure. The larger and more complicated our solution phase space, the more challenging is our optimization problem. Thus, a single, unique method of representing any given cell in terms of completely independent variables (along with implementations of the structure creation and variation operators which acted on and maintained this representation) is desirable. In practice, no such representation is known, and we are forced to search over a solution space which is highly redundant.

Additionally, the mating operation acts directly on the representation of structures. If, for example, one parent's unit cell is nearly cubic while the other's is very oblique, the real space distances between atoms may be greatly distorted in the child. Two crystals which are very similar physically could create an offspring which has little in common with either parent if their representations are sufficiently different. The naïve mating operation is most successful when the parents have similar lattice parameters, and more generally, when they are represented similarly in the computer.

In practice, both of these issues may be ameliorated by standardizing the representation of structures as they are created. This is done first by imposing hard constraints described in Section 2.3.4. Secondly, we use the Niggli cell reduction algorithm to reduce redundancy and attempt similar representation of cells.[87] The Niggli cell for any lattice is unique and has the shortest possible

lattice lengths. We transform all structures into this representation during the development stage. In practice, this results in cells which are as cubic as possible and greatly improves the efficacy of the mating operation and success of the algorithm.

We also often avoid performing operations such as random translations or rotations before mating. While these techniques can help to replicate favorable local structural motifs to different parts of the unit cell, they also force the algorithm to explore multiple redundant regions of the solution phase space. Additionally, we have found that the operation is more likely to result in successful offspring if its parents have approximately the same number of atoms. Thus, we have modified the mating operator to take a supercell of one of the parents if it can make the two parents closer to the same size by doing so.

There is an alternate approach to avoiding the poor results often obtained from mating unlike structures. Instead of trying to force the whole population to look alike, we could preferentially mate similar structures. This has a biological analogy: two dogs are more likely to produce successful offspring than a dog and a cat. It also makes sense from a configuration-space search point of view. If the algorithm is thought of as sampling the whole space and then zeroing in on several promising regions in parallel, it makes more sense to mate structures from the same promising regions in order to explore that region more thoroughly. This extension to our algorithm will be explored in future work.

2.3.6 Duplicate structure identification

The most computationally demanding part of structure searches is usually the energy evaluation. We would like to minimize the number of these computations. It is especially undesirable to run multiple energy calculations on a single structure. Unfortunately, it is possible to encounter duplicate structures quite often in a naïve implementation of the evolutionary search. If a pair of structures mate more than once, they are likely to create similar offspring. If the set of best structures does not change from generation to generation due to promotion, the set of parents, and thus the resulting set of children, can be very similar also. In addition, as the generation as a whole converges to the global minimum, all of the organisms are likely to become more similar. What is worse: once a couple of low energy, often-selected organisms are in the population, they can frequently reproduce and similar structures can effectively fill up the next generations. This leads to premature convergence which is in practice indistinguishable from convergence to a correct global minimum.

Maintaining genetic diversity is thus necessary to prevent unnecessary computations and premature convergence. Some authors have established a "δ-value" rule which disallows more than one organism in any generation with energies very close to one another.[37] However, the size of the interval is fairly arbitrary and system-dependent, and this method may lead to many false positive matches. The solution is to compare potentially duplicate structures more directly. Specifically, our algorithm keeps two separate lists of structures it has seen and explicitly checks against each of them when creating new structures. The first, a "per-generation" list, holds the relaxed, developed structures which are members of the current generation. Similarly, a "whole-population" list

holds all structures the algorithm has seen, both relaxed and unrelaxed.

The goal of the whole-population list is to prevent redundant energy calculations. If a new, unrelaxed structure is found to already exist on this list, it is discarded. The per-generation list serves to prevent premature convergence due to multiple occurrences of a single structure in a generation. It works similarly to the first list and, indeed, does not need to keep track of unrelaxed structures since the whole-population does that. Only organisms which are added to the generation are added to the per-generation list. Matches against the per-generation list can cause us to discard structures even after they have been relaxed. Throwing away the results of total energy calculations may seem unfortunate, but it is necessary to avoid both premature convergence and future redundant calculations. Seeing identical structures created in future generations is much less common if we take these steps to maintain genetic diversity in the first place.

Other authors have taken slightly different approaches to solving this problem.[96] Lonie *et al.* describe their approach to structure comparison for this application in Ref. [94]. Abraham and Probert enforce genetic diversity by adding a term to the fitness function which penalizes structures which are too similar to the lowest-energy structure with the same number of atoms according to a measure of similarity based on the structures' spherically-averaged scattering intensities.[4]

2.3.7 Composition-space search

We are often interested in how the structure of materials changes over a range of compositions. Our extension of the method to search over composition space, i.e. to explore a whole phase diagram at once, is based on that of Trimarchi *et al.*[154] and requires a modification of the single-composition method. Single-composition heuristic optimization methods search for stable structures by minimizing the total energy per atom. Here, we change the objective function to be the formation energy with respect to the structures on the currently-known convex hull. The convex hull is built and updated over the course of the algorithm – essentially, we add new energy values as the algorithm generates new structures and regenerate the convex hull curve when necessary.

For clarity, this process is illustrated for the case of a binary system in Fig. 2.5. In Fig. 2.5(a) we randomly generate our initial population and plot their energy versus composition. A convex hull is generated, and potential parents are evaluated according to their vertical distance from the hull, as illustrated in Fig. 2.5(b). The structures on the hull are the best, followed by the ones closest to it, and so on. In Fig. 2.5(c), we use those parents to create the next generation of structures. If structures with energies below the current convex hull are discovered, we need to redraw the convex hull as shown in Fig. 2.5(d) in preparation for the next generation. This iterative process proceeds until we have sufficiently searched the space. This algorithm generalizes straightforwardly to systems with more than two components.

A few practical issues arise in the extension of the genetic operators from the single-composition search to the phase diagram search. A convex hull must include at least some elemental reference states. We often draw the elemen-

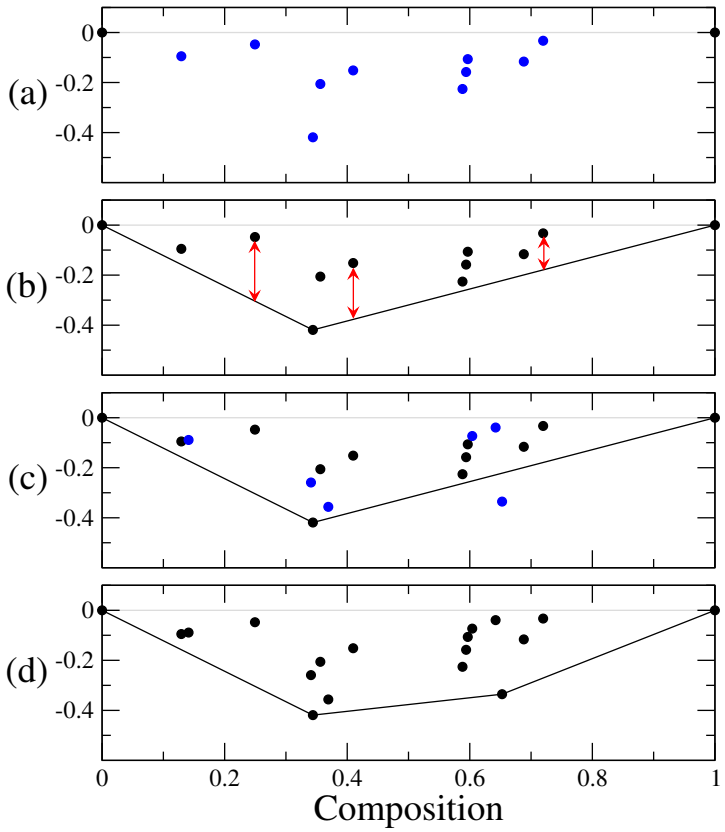


Figure 2.5: Illustration of the phase diagram searching procedure. See the text for an explanation.

tal structures from the literature or perform preliminary single-composition searches to determine them. If we are confident in these structures, we may even restrict the algorithm to only search over multi-component candidate structures. Also, as mentioned previously, the promotion operator usually acts in the case of the phase diagram search to maintain all the structures on the convex hull from one generation to the next.

The primary challenge remaining after a naïve implementation of the formation energy minimization algorithm has to do with searching over the composition degrees of freedom. Both the mating variation and a certain mutation-type variations can introduce solutions with a new composition, that is, they can cre-

ate children with compositions which differ from those of the parents. However, if we rely on these variations only to explore other compositions, it is not clear that we are effectively leveraging the power of the evolutionary process to optimize this parameter. For the evolutionary algorithm to be effective, the traits of solutions which are important for the fitness must be heritable.

Unfortunately, parents with good compositions do not tend to have children with good compositions, for two major reasons. Small changes in composition usually have large energy penalties. In this respect, the composition variable differs from the structural degrees of freedom. Assuming our energy calculation incorporates local optimization, the energy landscape with respect to the structural degrees of freedom is generally fairly smooth. Lyakhov *et al.* show that materials' energy landscapes often consist of "funnels" composed of fairly low-lying structures surrounding even lower minima.[96] There are no such funnels surrounding minima in the formation energy landscape with respect to the composition parameter, and there is no "relaxation"-type operation for these degrees of freedom since the composition parameters are not continuous. For this reason, relying on the mutation and mating operations designed for the single-composition search to vary the composition parameter may be thought of as essentially performing a random search over that parameter.

Furthermore, there is reason to believe that the passive search over composition space could sometimes be even worse than a true random sampling. Consider the case of parent structures that contain small numbers of atoms. For example, in early generations of our search of the Si-H system under pressure we found low-enthalpy pure-H and pure-Si structures with 1 and 4 atoms, respectively. If these two structures were to act as parents, the child could take

on only compositions H_iSi_j where i and j are integers such that $0 \leq i \leq 1$ and $0 \leq j \leq 4$. Early in a search, it could be quite likely for the elemental ground states to be often chosen as parents, and here it is not even possible to sample any of the important hydrogen-rich side of the phase diagram corresponding to silane, SiH_4! To prevent this problem, we have modified the mating operator to take a supercell of one of the parents with some probability before making children.

Furthermore, the method has a tendency to favor the sampling of intermediate compositions. For example, in order to sample an elemental phase, the algorithm will usually need to choose two parents with compositions which are equal, or at least very close, to that elemental composition. This has a very low probability and the sampling of intermediate compositions is much more likely.

We suggest two potential solutions to this issue. The first involves modifying the selection logic. Instead of determining the selection probability of each structure based solely on the value of the objective function, i.e. the distance above the convex hull, we modify it so that it is more likely for structures with similar compositions to be parents. As mentioned previously, this idea has roots in the biological analogy. In effect, structures at different compositions would evolve somewhat separately with a lesser degree of intermixing.

One might worry that this idea could lead to less effective sampling of composition space by forcing the search to focus too much at compositions around the best structures of the first generation. This danger could be ameliorated by repeating the broad sampling of the solution phase space after each time we identify a new point on the convex hull. Since the objective function changes with the convex hull, it makes sense to repeat this broad sampling of the so-

lution phase space, but this would have to be balanced with the maintaining of knowledge the algorithm has already gained about the system. In practice, however, there is sufficient randomness in the results of the variation operations, and the elemental phases are always fairly likely to be chosen as parents, so the algorithm is unlikely to be trapped in a local minimum with respect to the composition degrees of freedom.

A second way of addressing the tendency to favor sampling of intermediate compositions involves a reformulation of the algorithm. Its implementation in the GASP code is still in progress, but we describe it here. The idea is to perform a series of genetic algorithm runs, each on constrained segments of the composition space. The segments themselves are determined recursively as in the following pseudocode.

```
recursiveSearch(minComp, maxComp) {
  doGARunWithCompositionBetween(minComp, MaxComp);
  if (run improved convex hull) {
    Composition c = composition ...
        of best new stable structure;
    recursiveSearch(minComp, c);
    recursiveSearch(c, maxComp);
  }
}
```

This algorithm takes two endpoints in the binary composition space. It uses the genetic algorithm to search for stable structures with composition within the given range, including the endpoints. Since the algorithm can generate new

structures with the composition of the endpoints, the search has the potential to improve previously found structures on the convex hull. When the algorithm finds a new stable structure, it further divides the range of compositions it was given and makes a recursive call on each of the sub-ranges. This algorithm ensures that all of the composition space is sufficiently sampled. It also gracefully enforces mating between structures of similar composition at the lower levels of recursion.

2.3.8 Varying N

The problem of identifying the correct number of atoms in the cell is similar to the problem of searching over composition. The number of atoms in a child does depend strongly on the number of atoms in its parents, but since small changes in this parameter lead to large, un-relaxable changes in the energy, the evolutionary process is not particularly efficient for optimizing this parameter.

In some GA designs, the number of atoms in a unit cell is held fixed over all structures in the search. If a newly generated structure does not contain the required number of atoms, it is either discarded or modified. This approach is non-ideal for several reasons. First, successful *ab-initio* prediction in this case requires either a lucky guess at the correct N (or a multiple thereof), or many runs at different N. This second option leads to much duplication of work in the sense that good local structural motifs must be rediscovered by each run. For this reason, we find it best to allow the number of atoms in a unit cell, N, to vary over the course of the algorithm.

Secondly, if N is fixed, one would like to choose it to be relatively large so

that nontrivial structures can be represented. Lyakhov et al. describe some of the challenges associated with the random generation of an initial population of large structures. They show that the random generation of a large (on the order of 100-atom) cell has a very strong tendency to create a disordered, high-energy structure, and that an initial population of these offers "virtually no route toward the ordered ground state".[97] This problem is addressed by instead creating small random cells and then generating supercells of these to get initial structures of the appropriate size. This method is found to be successful in lowering the energy per atom of the randomly-generated initial population.

However, one issue with this approach is that the energy calculations are much more expensive than necessary. Even if the initial evaluations of these supercells can be simplified due to their high symmetry, the first generation of children will not share that property. More critical than the individual evaluations, however, is the complication of the search process itself. The number of local minima in the energy landscape is known to grow quickly with the number of atoms. So, forcing the algorithm to search over a part of phase space which is highly redundant in that it does not correspond to the simplest representation of a structure exacerbates the search problem.

Our approach to searching over structures with different N is as follows. We initialize the search with a population of relatively small structures, e.g. containing up to 8 atoms. Then, we allow and encourage the average structure size to grow over the course of a run. Pressure to increase system size is imposed in a very straightforward manner during the mating algorithm. As mentioned previously, for the mating operation we generally grow the smaller one of the two parents in order to make both parents similar in size. In addition, we define

some nonzero probability of doubling the number of atoms in one of the parent cells prior to this. In this way, we start out with simple structures, but then allow the algorithm to encounter larger, more complex ones after it has gathered more information about the structural motifs favored by the system. By varying this doubling probability, we can tune the speed with which the average cell size in the population increases.

2.4 Algorithm evaluation

We are interested in quantifying the success of the algorithm after it has used some number of energy calculations. In the case of a search performed at a single composition, we can simply look at the energy of the best structure encountered so far. However, since the method is stochastic in nature, we actually need to describe the distribution of this energy. The distribution can be found approximately by repeating the GA run many times with identical parameters: N_{runs} times, in particular. Each of these runs consists of N_{calcs} total energy calculations. For each of the N_{runs} genetic algorithm runs, we keep track of the best energy per atom encountered after i energy calculations with i ranging from 1 to N_{calcs}. Then, we plot the 10^{th} percentile, the median, and the 90^{th} percentiles of these best energies as a function of the number of energy calculations i. We will refer to these plots as performance distributions.

Performance distributions are closely related to the Hartke plot construction.[93, 62] Data for a Harkte plot is generated in the same way, and then the lowest best energy, the highest best energy, and the average best energy are plotted. However, the lowest and highest best energies encountered are outliers, and in

practice they depend strongly on the choice of N_{runs}. The 10^{th} and 90^{th} percentiles offer a better characterization of the distribution of results.

In the case of the optimization at a single composition, the lowest energy encountered after some number calculations gives a natural indication of the quality of the algorithm's results at that point. In the case of the phase diagram search, no single energy fills this role since the energies over all of composition space are relevant. That is, no single structure is the current best solution – the whole phase diagram is the solution. To quantify the success of the phase diagram search, we evaluate the area or volume of the convex hull itself with respect to fixed elemental phases. This quantity may not be directly physically meaningful, but it will be at a maximum when we have the ground state phase diagram. The closer the volume of the convex hull is to that maximum, the closer the phase diagram is in some average sense to the right answer. For this reason, the single-composition performance distribution curves decrease with the number of energy evaluations while they increase for the phase diagram search.

A performance distribution gives an indication of how well the algorithm could do, how poorly it could do, and how well it does on average, after investing a certain number of energy calculations. We will compare the performance of the algorithm for various parameterizations with a standard random search algorithm which involves relaxation from starting structures whose lattice parameters and atomic locations are randomly chosen within some constraints.[149] These structural constraints are the same as used for the GA searches, and the random structure generation logic is the same as is used to generate the GA's initial population. We will fix $N_{\text{runs}} = 100$, and N_{calcs} is set to

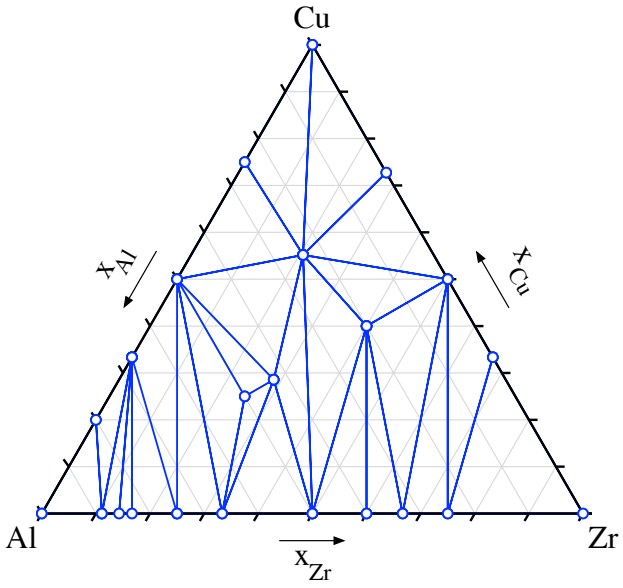


Figure 2.6: Zr-Cu-Al ternary phase diagram predicted by the EAM potentials.

500 for single-composition runs and 2000 for the phase diagram runs. Each of these runs cost on the order of several CPU hours on a local workstation.

Additionally, we are interested in comparing the phase diagram prediction method described above to a naïve alternative. The alternative is to compile a convex hull by performing a single composition search at many different compositions. To this end, we will compare the number of energy evaluations necessary to solve the phase diagram search to the number necessary to solve the single composition search times the number of single composition searches necessary to fill out a phase diagram.

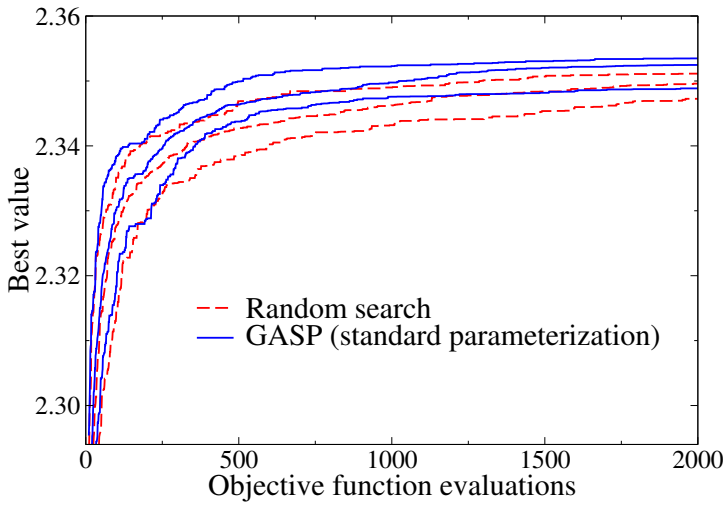


Figure 2.7: Performance distributions describing the success of the random searching method and a standard parameterization of the GA in finding the ternary phase diagram. The three sets of lines for each structure search methods correspond to the 90^{th} percentile, 10^{th} percentile, and median of the best volume of the convex hull for 100 instances of the composition search algorithm.

2.4.1 Composition space search with empirical potential

Figure 2.6 shows the convex hull of the EAM potential. The system exhibits many binary ground states and has ternary ground states with compositions Zr_2Cu_2Al, $Zr_2Cu_2Al_3$, $ZrCuAl_2$, and $Zr_6Cu_{16}Al_7$. Most of these were identified by the GA, but the last is an experimentally-observed structure. This structure has a 116 atom unit cell (cells of this size were excluded by the hard constraints during the GA search) and lies below the convex hull identified by the GA. On the other hand, none of the new structures identified by the GA search of the EAM energy landscape turned out to be more stable than the experimentally-known structures according to DFT which is, of course, as expected.

Figure 2.7 shows two performance distributions that compare a random

search and the standard parameterization of our GA. The algorithm used a generation size of 50 and ran for an average of 23.5 generations until reaching the target number of objective function evaluations. Focusing on the median best energies in the performance distributions, we observe a similar performance for the GA and the random search during the early stages of the search, i.e. when the GA essentially is a random search. Then, after only about 200 energy evaluations the evolutionary process kicks in, the population starts to gain more information about the problem, and the GA pulls ahead. The benefits of the GA become obvious after a few hundred energy evaluations. After about 1000 energy calculations, the 10^{th} percentile of GA results is comparable to the median of the random search, and the median GA result performs significantly better than the 90^{th} percentile random search.

This observation shows that the GA significantly outperforms a naïve random search. Still, the convex hull does not converge on average after 2,000 total energy calculations. This illustrates that ternary phase diagram prediction is a hard problem and remains unsolved. Increasing the dimensionality of the solution phase space, especially by adding unrelaxable degrees of freedom like increasing the number of atoms or components, greatly increases the difficulty of the prediction problem, as discussed previously. That said, the performance distribution curves are still improving after 2000 calculations, and running longer may certainly be reasonable for some applications.

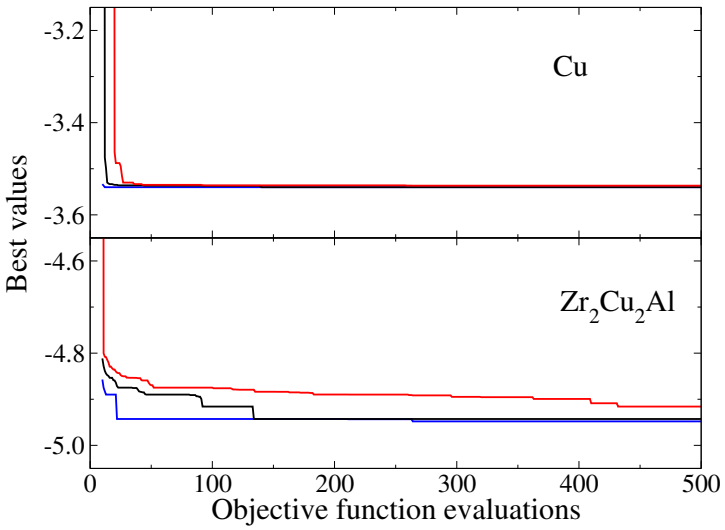


Figure 2.8: Performance distributions describing the success of a standard parameterization of the GA in finding the ground state at two particular compositions, elemental Cu and ternary Zr_2Cu_2Al. The three sets of lines for each composition search methods correspond to the 90^{th} percentile, 10^{th} percentile, and median of the best energies for 100 instances of the genetic algorithm structure search.

2.4.2 Evaluation of the phase diagram search

To evaluate the performance of the phase diagram searching method, we compare the runtime of our method with that of sampling compositions independently. Figure 2.8 shows the performance distribution of the GA searching at two specific compositions, elemental Cu and ternary Zr_2Cu_2Al. As expected, we observe that the speed with which the algorithm finds the ground state depends strongly on the complexity of the system. To find the fcc ground state structure of Cu requires on average about 50 energy evaluations, while for Zr_2Cu_2Al more than 500 energy evaluations are necessary.

Assume that it takes 300 evaluations to determine the ground state at a single composition. We furthermore assume a set of allowed compositions for

$Zr_iCu_jAl_k$ such that the integers i, j, and k can have values $0 \leq i, j, k \leq 8$. This results in a total of 571 unique compositions. Searching each of these compositions individually would require $300 \cdot 571 \approx 1.7 \cdot 10^5$ energy evaluations, far more than required by the method of Trimarchi.

Of course, this estimate might be a little misleading. On the one hand, the naïve method is certainly likely to search over some individual compositions better than the phase diagram searching method. On the other hand, the composition search at least has the possibility of generating structures at compositions which would be skipped entirely when sampling an arbitrary set of discrete compositions. Our observations from the performance distribution in Fig. 2.7 indicate that the GA goes through a learning phase during which it acts essentially as a random search, but after this phase, it is able to significantly outperform the random search. Since information about good local structural motifs is something that can be reused at different compositions, repeating this learning phase at every individual composition represents significant duplication of effort.

2.5 GASP software design

The GASP code which implements this genetic algorithm is meant to be a robust, user-friendly software package appropriate for use by a wide variety of researchers. However, it is also a research code in that its requirements are constantly changing to accomodate new applications and ideas. The software needs to be engineered to support both of these requirements – flexibility and robustness. Often these goals can be in conflict, since frequenty changes to a program

can lead to bugs. GASP is implemented in Java, and modern object oriented programming principles have been applied to allow for flexibility with respect to the objective function and the geometry of the system being studied. This allows us to easily solve new types of problems. Here we will describe the software design that makes it possible.

At the heart of the algorithm is a loop which essentially just generates new structures and then evaluates how good they are. The evaluation process needs to be somewhat different depending on the sort of system we are considering. In the simplest case, we are searching for the most stable crystal structure at a particular composition. Here the right quantity to minimize is the energy per atom. So, we can just take our crystal and pass it to one of the many energy codes with which we are interfaced and divide the resulting energy by the number of atoms in the cell.

Next, we might want to search for the most stable atomic cluster or molecule at a particular given composition. In this case, the core of the algorithm works on a cluster of atoms as if they were as a group of atoms in a crystal's unit cell. By and large, we use the same energy codes to evaluate them, but many of them assume periodic boundary conditions. Thus, to search for atomic structures is that, prior to the energy computation, the group of atoms needs to be packed into a large box so that there is enough vacuum between the cluster and its periodic images. This packing (and subsequent unpacking) is a small tweak to the evaluation process, and then we want to leverage all of the other logic associated with the structure generation and energy evaluation. In other words, we pack the structure into a large box, use the pre-existing objective function logic to evaluate it, and then unpack it again. A similar packing process is used

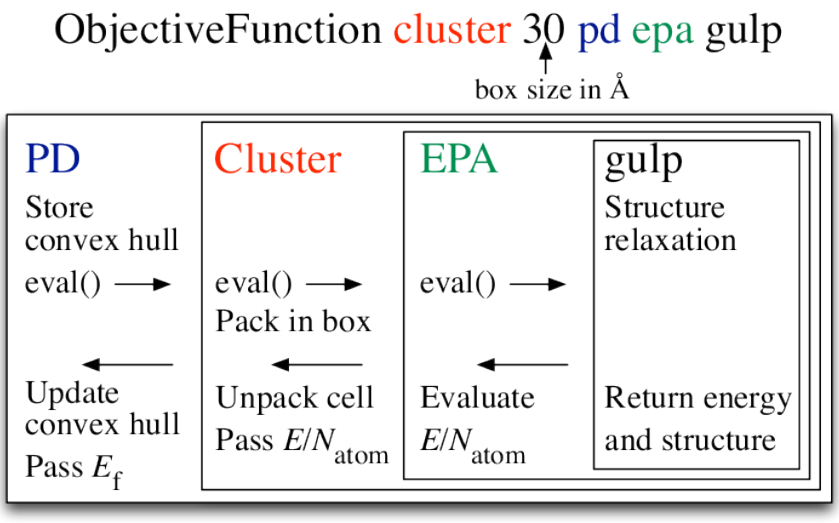


Figure 2.9: GASP's modular design allows us to easily implement new components and chain them together to achieve flexible run-time behavior and study many different systems, geometries, thermodynamics, and Hamiltonians.

to study 2D materials and surfaces.

We also want to support search over a phase diagram, allowing varying composition. In this case we need to evaluate new structures' formation energy with respect to all of the other structures the algorithm has seen so far in order to construct the convex hull. The module which evaluates new structures in a phase diagram search needs to keep a memory of the past structures, call the standard energy routines on the new structure, and then calculate a formation energy.

Finally, we often want to implement new capabilities and combine them in many different ways. For example, suppose we want to find the phase diagram of all atomic clusters according to a Hamiltonian described by the GULP code. This involves the same interface with the GULP code as we use everywhere else, but with a couple of tweaks. First we need to pack the atomic cluster in a large box, then we call the normal energy per atom GULP interface, and lastly we

need to transform that into a formation energy.

The challenge here is to combine all of these functionalities, allowing us to choose any combination of energy codes, geometries, thermodynamic ensembles, etc, and to do it in such as way as to

1. avoid code and logic duplication
2. let us combine the different components in any way we want
3. maintain sufficiently loose coupling between the different components to allow them to function and be extended independently

For example, with respect to the last point, it would not do for all of the geometry modules to have to know about all the different energy codes we might want to use since that leads to inflexibility and many points of failure which would very often be problematic in a research code where we regularly introduce new, unforeseen functionality.

A design pattern is a reusable solution to a commonly occurring software design problem, and our solution to this problem takes the form of the Decorator pattern.[53] It can essentially be thought of as follows. We define an ObjectiveFunction interface which has an evaluate() method. Whenever the main algorithm deals with any of these structure-evaluating modules, all it needs to know is that its evaluator implements the ObjectiveFunction interface, and thus it can be used to evaluate(). All of the modules we described above (for example, SurfaceObjFcn, ClusterObjFcn, EnergyPerAtom, etc.) adhere to this interface, and thus implement an evaluate() method. So, the rest of the code could be dealing with any of these when it goes to evaluate a new structure.

EnergyPerAtom, however, is the most basic and works with all of the energy codes to actually get a number. The rest of the modules are essentially decorators: they make a small tweak to the functionality and then use another ObjectiveFunction to finish the job. For example, if a ClusterObjFunction is asked to evaluate a structure, all it does is to pack that structure in a big box, call the evaluate method of another function, and then when that's completed, removes the vacuum padding from the relaxed structure. That lower-level evaluator could be an EnergyPerAtom, but it could another one such as the phase diagram module which itself uses another module under the covers, and our ClusterObjFunction does not need to know or care.

In this way, we can easily use a SurfaceObjFcn which itself uses an EnergyPerAtom which happens to interface to GULP under the covers. Or, we can use a ClusterObjFcn which use the phase diagram/formation energy evaluator which uses an EnergyPerAtom which happens to call VASP to do the actual energy calculation. And none of the modules knows what exactly other type of module it is dealing with – they just know that they are all ObjectiveFunctions. This design allows us to quickly extend the code to study other geometries or thermodynamics or to leverage new energy codes.

CHAPTER 3

PHASE DIAGRAM PREDICTION: BARIUM AND EUROPIUM UNDER PRESSURE

3.1 Barium

Much of the work in this section was previously described in Ref. [145].

Many alkali and alkali earth metals exhibit unexpected phase transitions at high pressures. Barium in particular was known previously to exhibit complex behavior, including a superconducting transition at high pressures, and so we selected it for further study. The experimentally-known phase transitions are illustrated in Figure 3.1. The material is bcc at ambient conditions. It is hcp at high pressures except for between about 13 and 45 GPa when the complex Ba-IV structure arises.

We performed a search for barium structures using the GASP software to search at seven pressures: 0, 10, 30, 60, 100, 200, and 300 GPa. These searches totalled approximately 1600 total energy calculations and identified approximately a dozen candidate structures which had low enthalpy at one or more pressures. Detailed enthalpy-vs-pressure curves for each of these are shown in Figure 3.2 with respect to the hcp phase. The experimentally-observed bcc-to-hcp transition is clearly reproduced. Additionally, this study identified a previously-unreported structure with prototype α-Sm, a close-packed structure with alternate stacking order. Our calculations predict that this structure is stable with respect to hcp at 35 GPa.

The Ba-IV phase has previously been experimentally characterized and is

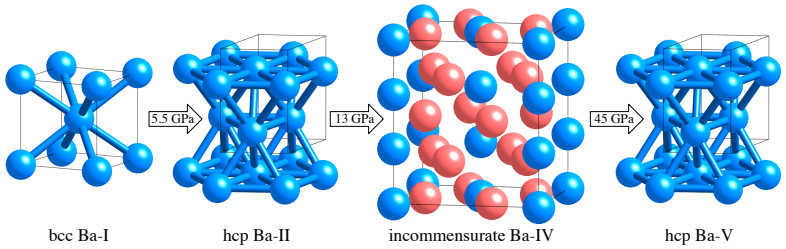


Figure 3.1: Experimental sequence of pressure-induced phase transformations of barium at room temperature.[103] The incommensurate Ba-IV phase is represented by the commensurate $^3/_2$ approximant with the host and guest sites represented by blue and red spheres, respectively.

very unusual.[103] It may be thought of as two coexisting crystal structures, a host cell which is bct and a guest cell which is fct. These two cells have identical lattice parameters in two dimensions but unequal c-axis lengths. Furthermore, the ratio of c_{host} to c_{guest} varies with pressure, but experimental work reports that at some pressures it is irrational. In this case, the overall material is aperiodic in one direction. This makes it impossible to model even the bulk material using periodic boundary conditions.

However, we can generate approximate Ba-IV structures with c-axis ratios near the experimentally-observed values. Some of these are illustrated in Figure 3.3. The 3/2 approximant has three guest and two host unit cells, distinguished by color in the figure. The experimentally measured c-axis ratios vary between 1.39 and 1.355 from 12-17 GPa. These approximants have c-axis ratio $^3/_2$, $^4/_3$, and $^{25}/_{18}$.

We calculate the enthalpy of each of these structures, and predict the ground state enthalpy and c-axis ratio as follows. At each pressure, we have three data points describing the enthalpy and c-axis ratio of our three approximants. We fit a quadratic to these points to find a function describing enthalpy as a function of

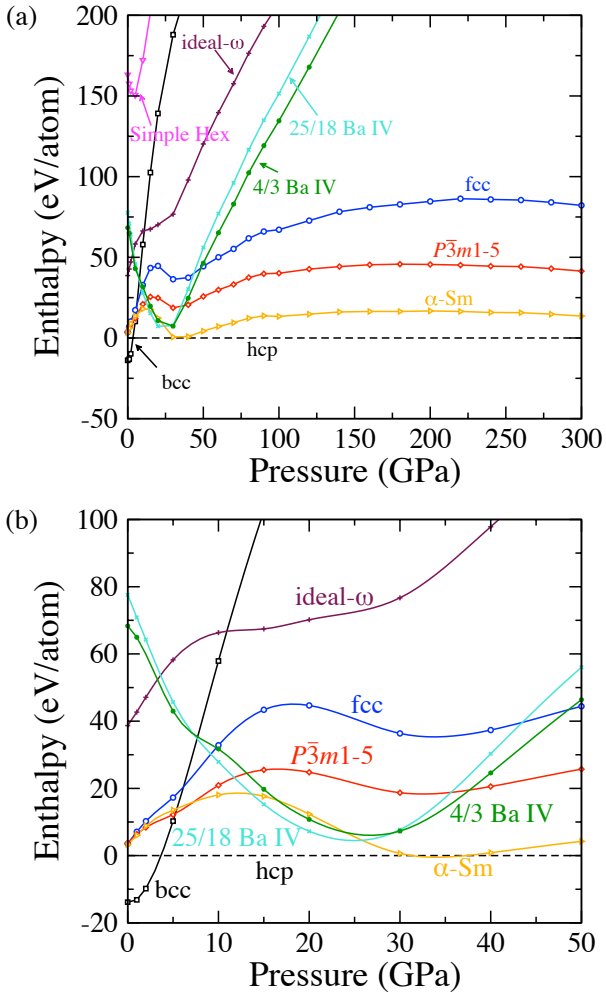


Figure 3.2: Enthalpy of barium structures as a function of pressure (a) over the entire range investigated, and (b) over the 0-50 GPa range. The enthalpies are given relative to that of the hcp phase which is the ground state at most pressures.

c-axis ratio. The true structure is that with the c-axis ratio corresponding to the lowest enthalpy. Minimizing this quadratic interpolation allows us to estimate the c-axis ratio and enthalpy of the ground state Ba-IV phase. Enthalpies as a function of pressure for the 3/2, 4/3, and 25/18 approximants and for the ground state Ba-IV phase estimated in this way are shown in Figure 3.4, and the corresponding ground state *c*-axis ratios are given in the inset.

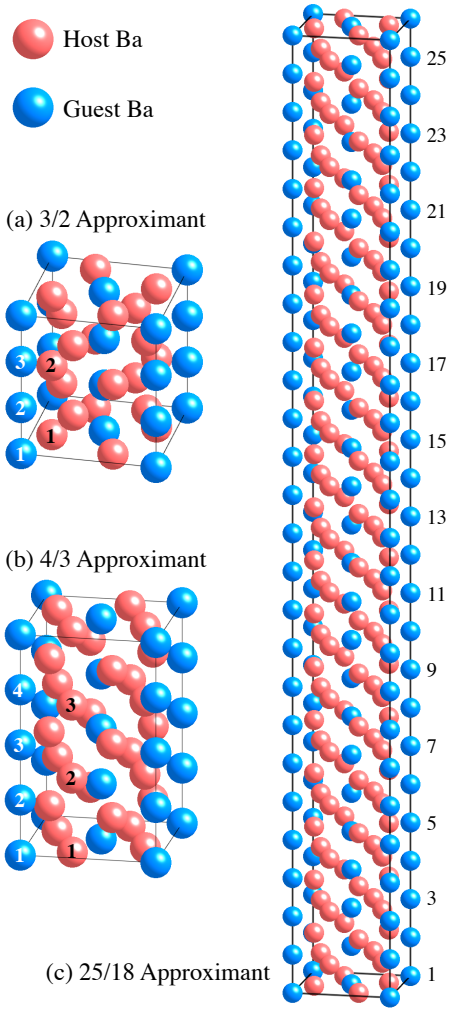


Figure 3.3: Approximant commensurate structures of the incommensurate Ba-IV phase, (a) 3/2, (b) 4/3, and (c) 25/18 (commensurate cells are named by [# of guest cells]/[# of host cells] convention). These structures are in space group $I4/mcm$. The 4/3 commensurate analog has 32 atoms in Wyckoff positions $4a$, $4c$, $8h(x = 0.35271)$, and $16l$ ($x = -0.35011$ and $z = -0.33503$). The structural parameters are for a pressure of 12.6 GPa.

3.2 Europium

Much of the work in this section was previously described in Ref. [17].

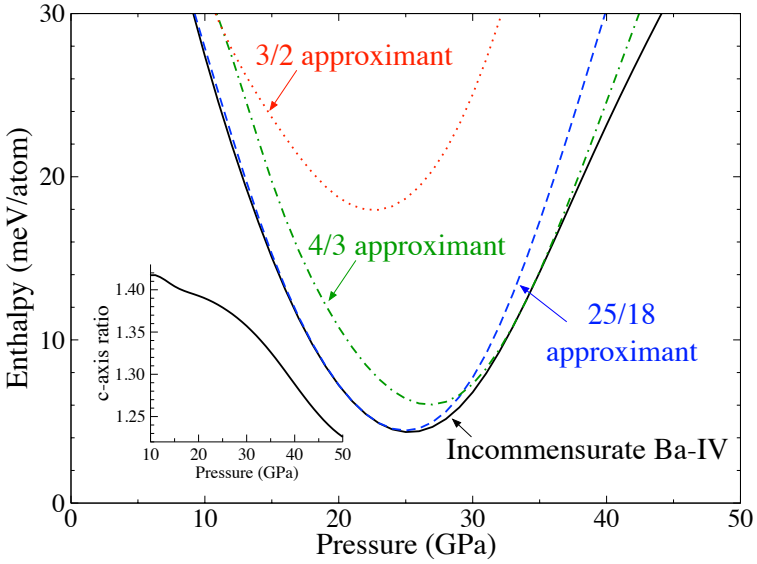


Figure 3.4: Enthalpy of the approximants to the incommensurate Ba IV phase and interpolation to the incommensurate Ba IV phase as a function of pressure. The inset shows the predicted change in $c_{\text{host}}/c_{\text{guest}}$ ratio as a function of pressure.

The GASP code was used in conjuction with experimental collaborators to solve for the structure of elemental europium up to 92 GPa.[17] Previous experiments only characterized the material up to 43 GPa[146, 61], and extending this work higher was important to check whether the material's superconducting transition at pressures above 80 GPa[38] is associated with a structural phase transition.

X-ray diffraction experiments performed in a diamond anvil cell are described in Ref. [17] and observed several europium phases up to 92 GPa. The structure is bcc up to 12 GPa. It is hcp from 12 to 18 GPa. Then, a mixed phase region is seen from 18 to 62 GPa. Finally, an orthorhombic structure arises around 66 GPa, and no further transitions are observed through the highest pressure applied. Solving the structure using the diffraction data involves checking the diffraction patterns of candidate structures against the experimental data until

a match is found. The GASP code was used to predict the phase diagram and identify these candidate structures.

We performed structure searches at pressures of 0, 20, 40, 60, and 80 GPa for structures with up to 30 atoms per unit cell. The search discovered a large number of candidate crystal phases with low enthalpies (bcc, fcc, hcp, Fdd2, Pnma, Fddd, Cc, Imm2, R-3m, C2/m, and C2/c). All of these structures have enthalpies within a range of 50 meV/atom. We calculate the enthalpy as a function of pressure for all of these trial structures. Figure 3.5 shows the enthalpies of the predicted ground state structures and their stability ranges. We find the bcc phase at low pressures and a transition to the hcp structure at 10 GPa. At a pressure of 16 GPa we predict a transformation to the C2/c structure, at 22 GPa to the Fdd2 structure, and at 34 GPa to the Pnma structure. The Pnma structure is nearly degenerate to the C2/c structure, and we predict that the C2/c phase is slightly lower in enthalpy above 46 GPa. However, these enthalpy differences are below the accuracy limits of current approximations of the exchange-correlation functional in DFT calculations.

As mentioned previously, the experimental diffraction data was matched to the hcp structure alone through 18 GPa. From there to around 28 GPa, the data was indexed to a mixture of hcp and an additional phase, and above 30 GPa, the transition proceeded more quickly. (This sluggish phase transition is consistent with small enthalpy differences between competing phases seen in the computational results.) The monoclinic C2/c was found to match this new phase and is hence assigned to be the post-hcp phase. Above 41 GPa, an additional phase appears in the diffraction data. This is the orthorhombic Pnma phase also identified by the genetic algorithm. Thus, the methodology implemented in the GASP

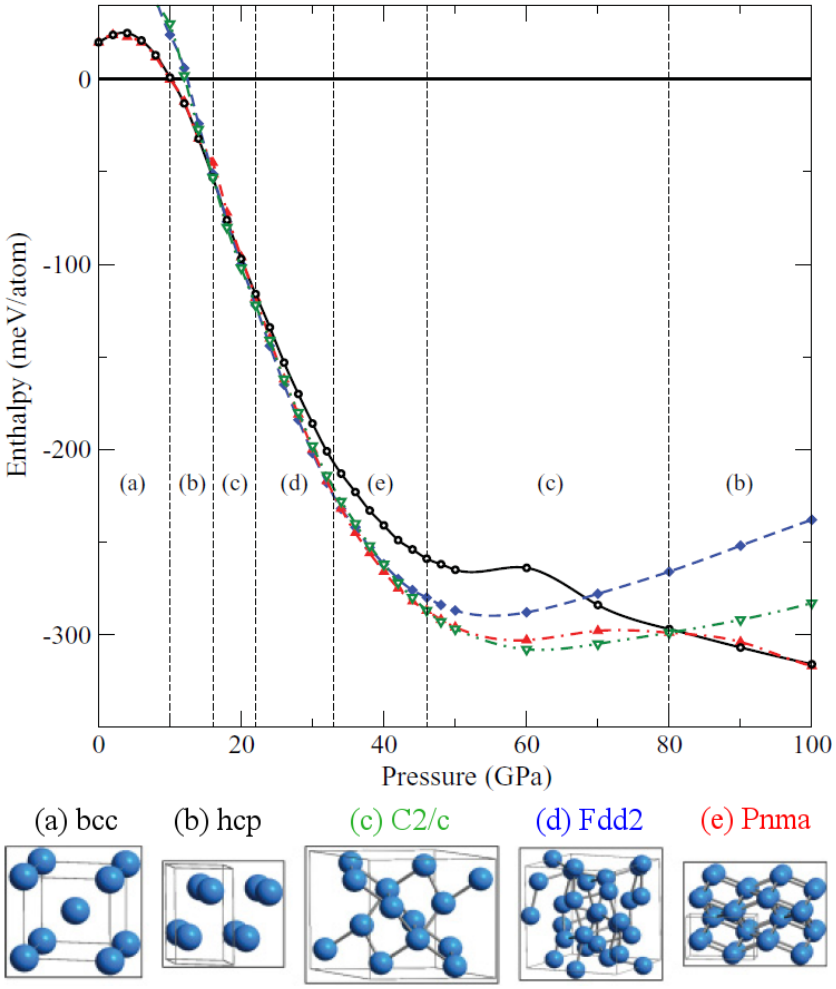


Figure 3.5: This figure from Ref. [17] shows enthalpies of candidate europium structures identified by the genetic algorithm search up to 100 GPa. The DFT results predict a structural sequence of bcc→hcp→C2/c→Fdd2→Pnma→C2/c→hcp, but the very small enthalpy differences between the competing phases are within the errors in the calculations.

code played a crucial role in interpretation of experimental characterization of elemental europium under pressure.

CHAPTER 4

CHARACTERIZATION OF LI-ION BATTERY MATERIALS

Much of the content of this chapter was previously published in Refs. [148] and [152].

4.1 The Li-Si system

4.1.1 Introduction

Lithium-ion batteries have proven their usefulness for electrochemical energy storage in portable electronic devices and electric vehicles. Current research efforts focus on improving the energy storage capacity, power density, and lifetime of Li-ion batteries.[48, 100] Among the various candidates for Li-ion anodes, silicon is the material with the highest known theoretical capacity. In fact, its measured capacity of 3500 mAh/g is about ten times the capacity of currently used graphitic anodes (372 mAh/g).[105, 26]

The lithiation of Si anodes is associated with a volume expansion of 300%. Silicon nanowires and nanotubes have been shown to accomodate this immense volume expansion and still provide reversible charge capacities which represent significant improvement over current devices.[58] When silicon anodes are lithiated at room temperature, they become amorphous and remain so under cycling for potentials above 70 meV vs. Li/Li^+. Full lithiation at potentials below 70 meV leads to the formation of crystalline $Li_{15}Si_4$.[89, 88, 105]

In this work we demonstrate that *ab-initio* structure searches can successfully

determine structural characteristics and energetics of battery electrodes, opening the possibility of the computational discovery of novel battery materials.

The core step in this process is an understanding of the atomic structure of the anode during the lithiation process. The equilibrium Li-Si phase diagram is well understood [110] – a number of Li-Si compounds are known and have been characterized experimentally. A comprehensive *ab-initio* study of the electronic structure, phonon spectra, and charge transfer characteristics in these structures was performed by Chevrier *et al.*[34] The five experimentally known phases are, from lowest to highest Li content, LiSi, $Li_{12}Si_7$, Li_7Si_3, $Li_{13}Si_4$, and $Li_{21}Si_5$. A phase with composition Li_2Si has been reported but lies about 40 meV above the ground state at 0 K according to the calculations described below.[109]

Most of these crystals have large unit cells, and kinetic limitations impede their formation during the charge/discharge cycle. In fact, the Li-Si anode material is known to become amorphous in a battery.[89, 88, 105] Thus, to study the energetics and the electrical potential of the system during the rapid lithiation and delithiation of the anode in a battery, we must go beyond the known equilibrium phase diagram and model the material system's amorphous and metastable crystalline phases. In addition to the experimental phase diagram, we consider small unit cell crystal structures from a genetic algorithm search and models of amorphous structures from *ab-initio* molecular dynamics simulations.

The calculated electrical potential profiles show that the amorphization of Si only slightly increases the anode potential, not significantly affecting the operating voltage of the battery. We compare the structure of the amorphous phase to the metastable crystal structures and find that these small-unit cell structures

approximate the energies of and the local structural environments found in the amorphous phases. The ability of small structural approximants to accurately represent the energies of larger systems as has also been noted for example in quasicrystalline systems, *cf.* Ref. [[64]].

As computational methods have become more accurate and their implementations increasingly efficient, they have taken some of the burden of finding and characterizing new materials off experimental work.[82] High-throughput techniques that make efficient use of pre-existing information have successfully been used to identify battery materials. However, these techniques are not applicable to systems which have not been well characterized previously. In the case of simple binary systems such as Li-Si, complete structural data is often available, while this is not so for many more complicated ternary and quaternary systems, *etc.* In these cases, *ab-initio* predictions using data-mining methods alone are not necessarily sufficient. Leveraging a structure prediction method can overcome this limitation. Additionally, the problem really requires phase diagram prediction, as opposed to predictions at a single composition. Genetic algorithms that search directly over structure and composition space can provide a solution to this problem. The approach presented in this work enables the *ab-initio* exploration of new materials systems which hold promise for use in batteries and the prediction of the characteristics of these batteries without experimental input.

4.1.2 Methods

Density-functional calculations

DFT calculations are performed with VASP,[85, 84] a density-functional code using a plane-wave basis and the projector-augmented wave method.[18, 86] For Si, the $3s$ and $3p$ valence states are treated explicitly. For Li, all three electrons are treated explicitly to accurately describe the interactions involving Li^+ cations in the Li-Si structures. A plane-wave energy cutoff of 350 eV and a k-point mesh of density 40 per $^{-1}$ ensure energy convergence to within 1 meV/atom. The Brillouin-zone integration is performed with the Methfessel-Paxton scheme as implemented in VASP and a smearing of 0.1 eV. For energy calculations and structural relaxations, the Perdew-Burke-Ernzerhof (PBE) generalized-gradient approximation to the exchange-correlation functional is used.[117]

We investigate the band structure of the LiSi phase. To overcome the bandgap problem of semi-local exchange-correlation functionals and to accurately determine the band gap of select structures, we employ the HSE06 hybrid functional. [69] Additionally, we use this method to optimize the geometry of the Li_5Si_2 phase.

Since the candidate structures generated by the genetic algorithm have little or no symmetry, they are unlikely to converge to stationary points that are not local minima of the potential energy surface. That is, no further local relaxations would lower their energy. For this reason, we refer to all of these local minima above the convex hull as metastable structures in this chapter.

Energy-lowering symmetry breaking in supercells is still potentially pos-

sible and would result in imaginary phonon frequencies. We use density-functional perturbation theory as implemented in VASP along with the Phonopy package[153] to calculate the phonon densities of states and verify the dynamical stability of the structures on the convex hull predicted by the structure search. 3x3x3 supercells of the structures' primitive cells were used to calculate the phonons.

Additionally, this method is used to estimate vibrational, including zero-point, contributions to the free energy at finite temperature for the small unit cell analogues identified by the structure search. This allows us to evaluate the effect of finite temperature on our predicted voltage curves. In particular, we predict the voltage curve at 690 K in order to facilitate comparison with the experimental data by Wen *et al*.[168]

Metastable Li-Si phases

Amorphous phases. Model amorphous structures for Si and Li_xSi_{1-x}, with x = 0, 0.3, 0.5, 0.6, 0.7 and 0.8, were generated to study the energetics of the system after it is amorphized through repeated cycling[1]. These were created by a melt-and-quench approach[133, 81] using constant-volume *ab-initio* molecular dynamics simulations of 100-atom simulation cells with a time step of 2 fs. The temperature was increased from an initial value of 100 K to a maximum value of 4000 K at a rate of 0.4 K/fs and then decreased to 25 K at the same rate. The final ionic coordinates and simulation cells in each case were then optimized within DFT using the conjugate gradients method. A reduced plane wave energy cutoff of 225 eV was employed during the molecular dynamics steps. Other

[1]The molecular dynamics calculations described here were performed by Clive Bealing, co-author on Ref. [148].

parameters used for the optimization step are as described above, including the density of the *k*-point mesh, although the error due to Γ-point-only sampling was found to be quite small, 17 meV/atom at most.

The initial atomic positions in the cells were generated randomly, at a density close to that of the known Li-Si crystal structure nearest in composition. At each stoichiometry, we generated 20 models in order to provide a reasonable sample of the amorphous material. To confirm the accuracy of our modeling approach for the energetics of the amorphous Li-Si structures, we re-annealed and quenched the four lowest-energy amorphous structures at the $Li_{0.6}Si_{0.4}$ stoichiometry, with a lower maximum temperature of 1500 K and a heating and cooling rate of 0.15 K/fs, and then re-optimized the structures. The re-annealing did not consistently produce amorphous model structures with lower energies than those in the original set. Instead, the average energy after re-annealing the lowest-energy structures increased towards the mean energy of the initial set. This test provides evidence that the annealing and quenching rate did not affect the energies of the amorphous structure models.

Previous authors have studied the insertion of lithium into silicon, but the mechanism of the transformation from the crystalline to amorphous states is still poorly understood.[80, 75, 32, 180, 162, 31] Chevrier *et al.* [31] modeled the change in volume, energy, and electrical potential of amorphous Si anodes under lithiation using small unit cells of 5 to 11 Si atoms and adding an increasing number of Li atoms into the largest interstitial spaces of the structure. Our approach extends this previous study to much larger simulation cells and differs in the method of generating the amorphous structure models. The similarities in the results of both models indicate that the volume and energy are not partic-

ularly sensitive to details of this methodology. However, further work would be useful in determining the effect of quench rates, cell size, and thermodynamic ensemble on the characteristics of the resulting amorphous structures.

Crystal structure search. We use our Genetic Algorithm for Structure and Phase Prediction (GASP)[150] code to explore the energy landscape and find low-energy, small unit cell crystal structures across the whole composition range of the Li-Si binary phase diagram. These small unit cell structures are expected to have share many of the local structural characteristics of the systems's large unit cell crystal structures and the amorphous ones which form during lithiation and delithiation. Additionally, since the most important energetic interactions in most materials are short range, these small cell structures provide good energetic proxies for the real structures, allowing *ab-initio* prediction of a cell's voltage characteristics. The local structural characteristics and energetics of the genetic algorithm structures, the experimentally known crystals, and the model amorphous structures are compared below.

The genetic algorithm generates successive groups (generations) of candidate structures in such a way as to utilize information gathered from early trials to make better guesses later on. In the initial generation, known elemental phases were supplied (although this is not necessary) and a variety of other structures are generated randomly, within some loose constraints on the volume per atom, inter-atomic distances, and crystal lattice parameters. New structures are evaluated according to their formation energy with respect to previously encountered structures – the lower the formation energy, the better the new solution.[155] The best of the solutions are probabilistically allowed to "reproduce" through one of several evolutionary operators which mimic biological

processes such as mutation and mating. Crystal structure mutation involves randomly perturbating the lattice and atomic coordinates, and the mating operation is based on the splicing-together of spatially-coherent chunks of parent structures.[3, 37, 57, 107, 74, 98, 23] Since good structures reproduce while bad ones do not, structural motifs which lead to low energy crystals are propogated in the population while less favorable features tend to die out. This process was repeated for about 35 generations, using about 1000 structural relaxations with DFT. The maximum number of atoms in the cell was fixed to 20 and the average number of atoms in the unit cells encountered by the algorithm was 8.5. A more complete description of the algorithm which discusses several of its other important features is forthcoming.

Convex hull construction.

The convex hull construction is used to determine the thermodynamic ground state phases and energies of the Li-Si system at any composition. The convex hull is constructed by first plotting a set of Li-Si compounds according to their formation energy as a function of their composition. The formation energies are defined relative to the pure elemental phases. The smallest convex shape bounding these points is known as the convex hull, and we are interested in its lowest-energy segments at each composition. The thermodynamically stable ground state phases are those which lie on the convex hull. For compositions that do not have a representative structure on the convex hull, the thermodynamic ground state is a phase mixture of the endpoints of the segment at that composition. Neglecting any interfacial energies, this mixed phase has an average energy corresponding to the value of the convex hull at that composition.

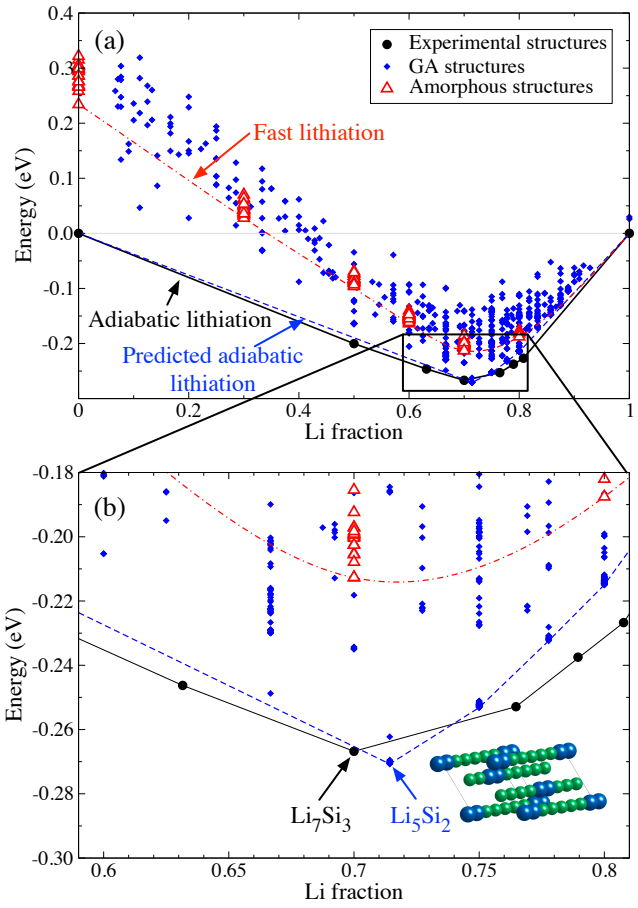


Figure 4.1: Formation energies of the crystalline, amorphous and ground state structures as a function of composition, (a) for the complete composition range from Si to Li and (b) for the region around $Li_{2.33}Si$.

Furthermore, the vertical distance between the formation energy of any given phase and the convex hull represents the energy between this phase and the energy of the thermodynamic ground state at that composition.

Thus, the convex hull provides an effective means to trace the energy of the system as a function of composition. Additionally, we will extend this useful concept to describe how the energetics of a system might change with respect to composition even in the case that it is not at equilibrium.

Electrical potential profiles

The electrical potential profiles show the open circuit voltage of a battery as a function of lithium content of the anode, from pure Si to the fully charged state $Li_{15}Si_4$.[34] These are generated from the formation energy of the phases. The electrical potential of the cell between two compositions $Li_{x_1}Si$ and $Li_{x_0}Si$ is given in Volts by

$$V = -\frac{G(\mathrm{Li}_{x_1}\mathrm{Si}) - G(\mathrm{Li}_{x_0}\mathrm{Si}) - (x_1 - x_0)G(\mathrm{Li})}{x_1 - x_0},$$

where $G(\mathrm{Li}_x\mathrm{Si})$ and $G(\mathrm{Li})$ indicate the Gibbs free energy of Li_xSi per Si atom and the Gibbs free energy of Li per Li atom, respectively, in electron-volts.[10] Neglecting entropic and zero-point contributions to the Gibbs free energy, we approximate G with the ground state internal energy.

4.1.3 Results and Discussion

Figure 4.1 shows the formation energies of the amorphous structures as well as those of the experimentally known phases. The figure also shows the approximately 700 metastable crystal structures that were encountered by the genetic algorithm. The formation energies of these structures are mostly above the convex hull of the experimental structures and hence metastable. However, a stable Li_5Si_2 phase is found by the genetic algorithm. The energies of the amorphous structures are above those of the lowest metastable crystal structures from the genetic algorithm search but comparable to their average energies. The four binary structures on the convex hull generated by the structure search are all dynamically stable as shown by their phonon densities of states in Fig. 4.2.

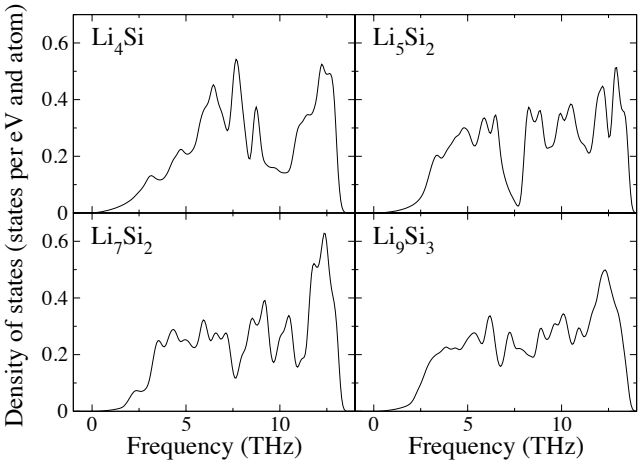


Figure 4.2: Phonon density of states of the four binary phases on the convex hull generated by the structure search show that these phases are dynamically stable.

Electrical potential of Li-Si

Several convex hulls are drawn in Fig. 4.1 to best represent what may be reasonable energies of the material during charging and discharging of the battery. The curve labeled *adiabatic lithiation* is composed of the experimentally known ground state structures. The convex hull labeled *predicted adiabatic lithiation* represents the ground state energetics of the system as predicted by the genetic algorithm. The *fast lithiation* convex hull is composed of our amorphous structures. This last curve is not a linear interpolation between points but actually a cubic spline fit to the lowest-energy amorphous structures at the compositions we sampled to represent how the energy of the system might change as a function of composition during rapid lithiation and delithiation of the Si anode.

Figure 4.3 shows the battery's electrical potential versus Li/Li^+ as a function of Li content for each of the three sequences. Comparing the calculated potential profile for the experimentally known phases to the data by Wen *et al.*[168]

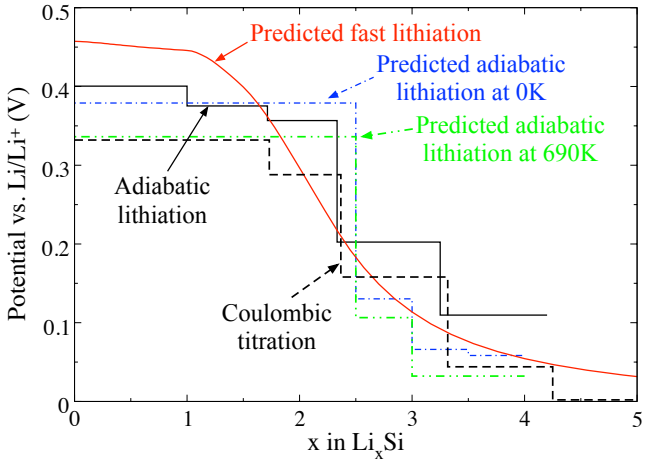


Figure 4.3: Electric potential of the Li-Si system as a function of Li content relative to the potential of Li/Li^+. The experimental result were obtained by Wen *et al.* by Coulombic titration. [168]

from Coulombic titration at 415°C illustrates the accuracy of this approach for this system. Previous DFT calculations by Chevrier *et al.*[33] showed both the local-density approximation and the generalized-gradient approximation (PBE) result in similar electrical potential profiles. Hence, we expect that the small differences between the calculated and measured potential profiles, which are of the order of 50 meV, are caused by the neglected entropic contributions to the Gibbs free energy. The voltage curve created with vibrational (including zero-point) contributions to the free energy agrees with the experimental curve extremely closely for the lithiation of Si up to Li_5Si_2, confirming this expectation.

We observe a similar behavior for the electrical potential as a function of Li content for each of the sequences of structures. Up to a Li content of about $Li_{2.5}Si$, the electrical potential is nearly constant. Further increase in the Li content decreases the electrical potential significantly from about 0.4 to 0.1 V. At low Li concentrations, the occurrence of the amorphous Si phase during fast and intermediate lithiation leads to a slight increase in the electrical potential of the Si

anode compared to the crystalline ground state phases. At high Li content, the potential may be reduced for intermediate lithiation rates compared to the fast and adiabatic lithitation due to the high stability of some of the small-unit cell structures.

To predict the maximum charge capacity of the Li-Si anode, we determine the maximum Li content of Li-Si possible while avoiding the formation of metallic Li. For the adiabatic lithiation, the maximum Li content is $Li_{4.4}Si$ corresponding to the experimental structure with the highest concentration of Li. For the fast lithiation, we estimate that the amorphous Li_4Si phase is the highest Li-containing structure. The resulting maximum charge capacity is 4200 mAh/g for the adiabatic lithiation and 3800 mAh/g for the fast case, closely corresponding to experimentally measured values ranging from 3500 to 3800 mAh/g.[26, 76] A conservative estimate of the charge capacity can be obtained from assuming lithiation up to a composition of $Li_{2.5}Si$. This would provide a nearly constant electrical potential and a still respectable charge capacity of 2400 mAh/g.

The $Li_{2.33}Si$ phase

The genetic algorithm encountered a noteworthy stable crystal structure, the Li_5Si_2 phase. Table 4.1 shows the crystallographic information of the Li_5Si_2 phase, obtained from relaxations with the HSE06 exchange-correlation functional.[69] Previous calculations for silicon have shown that the HSE06 functional accurately predicts the energies and structures of various phases and defects in agreement with experimental data and quantum Monte Carlo calculations and overcomes the tendency of the local-density approxima-

Table 4.1: Structure and vacancy formation energies of the Li_5Si_2 phase. The structure's space group is $R\bar{3}m$ (166) and its lattice parameters are $a = 4.383$ and $c = 17.837$. The structural parameters are obtained using the HSE06 functional. The vacancy formation energies are calculated using a 168-atom cell and the PBE functional. The vacancy formation energy differs dramatically between the Li and Si sites. The unusually low Li vacancy formation energy indicates a high Li vacancy concentration at room temperature and high Li mobility.

Site	x	y	z	E^f_{vac} [eV]
Li1 1(b)	0.0	0.0	0.5	0.015
Li2 2(c)	0.0	0.0	0.353	0.010
Li3 2(c)	0.0	0.0	0.210	0.049
Si 2(c)	0.0	0.0	0.067	1.11

tion and generalized-gradient approximation to overbind and underbind, respectively.[14, 66] For the Li_5Si_2 phase we find the same trend, the HSE06 functional predicts a volume of 14.13 ^{3}/atom, slightly smaller than the PBE prediction of 14.25 ^{3}/atom. Both potentials predict essentially the same Wyckoff positions.

Upon inspection, the Li_5Si_2 phase corresponds to the experimentally observed $Li_{2.33}Si$ phase.[159] It lies 8 meV/atom below the tie-line of the experimental structures and is stable against decomposition into Li_7Si_3 and $Li_{13}Si_4$. Von Schnering *et al.* determined the crystal structure of the $Li_{2.33}Si$ phase from single crystal X-ray diffraction and found a structure equivalent to Li_5Sn_2, the same one identified by our genetic algorithm but with partial occupation of 0.8 and 0.95 on the Li1 1(b) and Li2 2(c) Wyckoff sites, respectively, resulting in a composition of Li_7Si_3. Later, Barvik proposed two possible superstructures.[13] Chevrier *et al.* studied the partial site occupation of the Li_7Si_3 phase using these

superstructures and proposed a 60 atom representation of the $Li_{2.33}Si$ phase. [33] In this representation, every third Li1 1(b) site is vacant, yielding an occupancy of 0.66, and the Li2 2(c) sites are all fully occupied. Our energy calculations indicate that both structures, the fully ordered Li_5Si_2 and the 60 atom representation of the Li_7Si_3 structures with partial site occupation are representatives for the $Li_{2.33}Si$ phases that can be formed experimentally by annealing amorphous lithiated Si anodes with a composition a-Li_3Si.[33] The partial site occupation is expected to lead to an enhanced entropic stabilization of this $Li_{2.33}Si$ phase compared to the competing stoichiometric Li-Si line phases.

To elucidate the experimental observation of such an unusually high Li vacancy concentration in the $Li_{2.33}Si$ phase, we determine the vacancy formation energies for Li and Si on the various lattice sites. We perform this calculation for each of the 3 distinct Li sites and the Si site listed in Table 4.1. To reduce interactions between the vacancies we perform the calculations using a 168 atom simulation cell with a minimum distance between periodic images of the vacancies of 11.5 . For the chemical potential reference for the Li and the Si, we use the next most highly lithiated phase, $Li_{12}Si_7$.

Table 4.1 lists the resulting vacancy formation energies obtained using the PBE functional. The Li vacancies have unusually low formation energies of 10 to 49 meV. In contrast, the cost of forming a Si vacancy is 1.11 eV. The 60-atom representation of the $Li_{2.33}Si$ phase corresponds to the Li_5Si_2 structure with one vacancy per 15 Li atoms. This particular arrangement of Li vacancies leads to an average defect formation energy of 14 meV – very similar to our results for the 168 atom cell. These low Li vacancy formation energies will result in high vacancy concentrations at finite temperature. At such high vacancy con-

centrations, the interactions between vacancies becomes important and a more detailed analysis incorporating these effects would be necessary to predict the vacancy concentrations.

Electronic structure of LiSi

Electrical conductivity measurements of the LiSi, $Li_{12}Si_7$, and Li_7Si_3 phases have shown these to be small-gap semiconductors.[104] The experimental band gap for the LiSi structure was determined to be 0.06 eV.[143] However, due to the well-known DFT band gap problem, DFT calculations using semi-local approximations for the exchange-correlation functional do not reproduce the expected band structure and instead predict these phases to be metallic.[34] This discrepancy between experiment and theory along with the observed small density of states at the Fermi level for the semi-local functionals in the least-lithiated phases motivates further calculations using more accurate methods.

The primitive LiSi cell contains just 16 atoms, allowing for an efficient investigation of its electronic structure using the more computationally demanding hybrid functional approach.[69, 113] Hybrid functionals have been shown to be highly accurate in the prediction of semiconductor band gaps.[63, 68] The hybrid calculation was carried out using the HSE06 functional as implemented in VASP with a plane-wave energy cutoff of 350 eV, and a Monkhorst-Pack $8\times8\times8$ k-point mesh. The total energy is converged with respect to k-point density to within 5 meV per atom. The geometry was first re-relaxed, and then the electronic structure part of the calculation carried out.

The HSE calculations predict LiSi to be small-gap semiconductor with a band

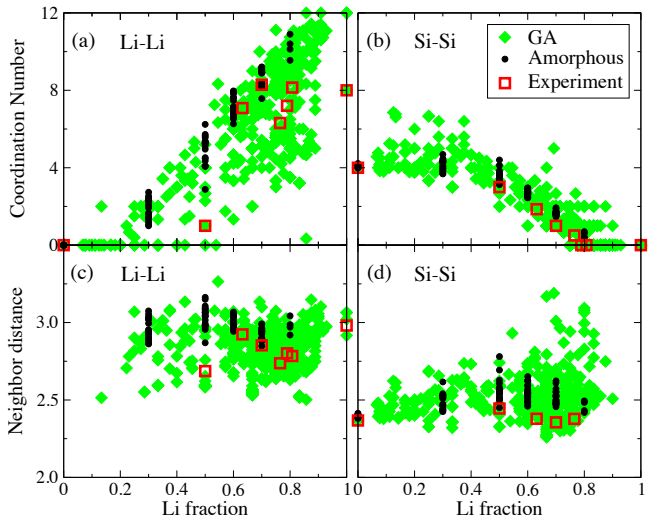


Figure 4.4: Comparison of the local structure of the amorphous, genetic algorithm, and experimental structures. Panels (a) and (b) show the coordination number for Li-Li and Si-Si neighbors, respectively, and panels (c) and (d) the average nearest neighbor distance in for Li-Li and Si-Si neighbors, respectively.

gap of 0.33 eV. This is somewhat larger than the experimentally determined value of 0.06 eV. It is noted, however, that the band gap for phosphorus black, which is isoelectronic and isostructural with the Si framework in LiSi, has also been determined as 0.33 eV.[143, 79]

Structure of metastable crystalline and amorphous phases

To explore how closely our genetic algorithm structures approximate the local structures of the amorphous phases and the experimentally known ground states, we examined the radial distribution functions of these structures. In particular, for each structure, we study the local environment of each type of atom, Li and Si. We determine the average number of nearest neighbors of the same type and the average distances to these neighbors.

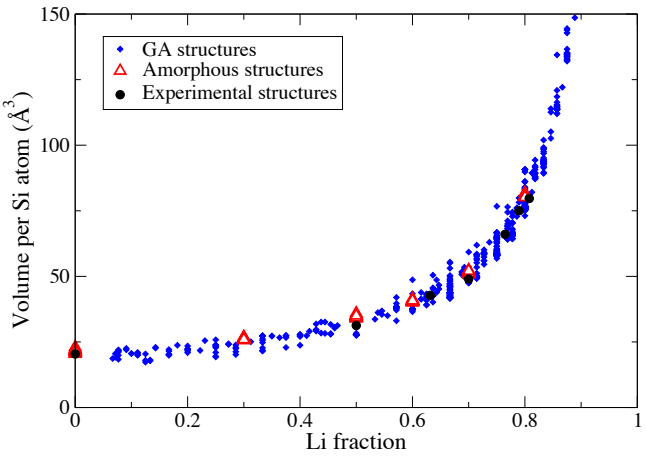


Figure 4.5: Change in volume of the anode as a function of Li content for the experimental ground state structures, the amorphous phases, and the genetic algorithm structures. The volume per Si atom increases from approximately 20 ^{3}/atom for pure Si to about 75-80 ^{3}/atom at a Li content of Li_4Si for all phases, a 400% volume expansion.

Figure 4.4 compares the local structure of the crystalline and amorphous phases. Results are shown for all amorphous and experimentally known structures and for structures encountered by the genetic algorithm whose formation energies are within 150 meV/atom of the convex hull. In general, the three types of structures exhibit similar trends for their coordination numbers and nearest neighbor distances as a function of composition. We observe that the Li atoms tend to be isolated up to a Li fraction of about 20%. At high Li concentrations, the Si atoms tend to form isolated dimers or trimers in all structures. On average, the Li and Si atoms in the ground state crystal structures exhibit somewhat lower coordination numbers and nearest neighbor distances than in the amorphous phases. The genetic algorithm structures, however, exhibit a large variety of local environments, some of which provide good approximants for the amorphous phases.

Figure 4.5 compares the change in volume for the three cases of lithiation

rate. The volume of the cathode expands from 20 ^{3} per Si atom for pure Si to about 75-80 ^{3} per Si atom at a Li content of 80% for all phases. The predicted volume expansion of 400% is consistent with experimental observations for Si nanoparticles and nanowires. [78, 26]

4.1.4 Conclusions

We described a methodology for determining the charge capacity, voltage profiles, and volume expansion of a Li-ion battery system solely by *ab-initio* methods without experimental input and demonstrated it for the technologically important Li-Si case. We performed a crystal structure search using the GASP code to identify low-energy small-unit cell crystalline phases. We also modeled the structure of amorphous Li-Si by *ab-initio* molecular dynamics simulations. It is known experimentally that the structures are dependent on the rates of lithiation and delithiation and we have suggested sequences of structures for adiabatic and fast rates of lithiation. For each case, we predicted the battery's electrical potential profiles as a function of capacity and determined the charge capacity of the anode. Our predictions of the battery's characteristics agree well with experimental observations.

We showed that the small-cell crystal structures share many local structural characteristics with the known ground state and model amorphous phases. This indicates that the phase diagram search methodology based on a genetic algorithm such as that implemented in the GASP code is useful for producing small-cell analogues which are good structural and energetic proxies for the material that might form when charge and discharge rates are too high for the system

to achieve the thermodynamic ground states. The structures encountered by such a search are thus useful for predicting many characteristics of the electrode and may furthermore be useful for a study of Li diffusion in the system. Similar structure search methods have been implemented in the USPEX[57] and XtalOpt[93] programs.

Additionally, we have predicted the stability of a binary structure, Li_5Si_2, that corresponds to the ground state structure of the experimentally observed $Li_{2.33}Si$ phase. We find an unusually low vacancy formation energy of Li in this phase that explains the observed change in stoichiometry under ambient conditions.

4.2 The Li-Ge system

The lithium-germanium materials system has potential uses in lithium-ion battery devices. An *ab-initio* genetic algorithm search of the system identifies a previously unreported member of the lithium-germanium binary phase diagram with composition Li_5Si_2 and space group $R\bar{3}m$ (166). We describe the structure of this new phase, investigate its mechanical and electronic properties, and discuss its relationship to other members of the lithium-germanium binary phase diagram.

4.3 Introduction

Most current lithium-ion batteries in modern devices contain a graphitic anode, since carbon is inexpensive, widely available, and has a low electrochemical potential relative to lithium metal. Graphite also has a layered structure, which allows lithium ions to easily be inserted between its layers. However, graphite has a relatively low theoretical capacity of 372 $mAhg^{-1}$ [147], and work to find a higher-capacity alternative is ongoing.

One candidate is Ge, which has a high theoretical gravimetric capacity of 1384 $mAhg^{-1}$ for the $Li_{3.75}Ge$ phase, as well as a high volumetric capacity of 7366 AhL^{-1} [90]. Ge has high cost relative to other materials such as silicon and carbon, but as an abundant element, its high cost is simply a result of the current lack of demand [60].

Recent experimental studies on the Li-Ge systems for batteries have focused on a variety of germanium nanostructures – for example, nanowires [147, 27], nanofilms [60], and nanoparticles coated with carbon shells [171]. The use of nanostructures result in an improved capacity and other properties when compared to the bulk form. This is due to the nanostructures' higher surface area and shorter diffusion lengths. Nanostructures are also effective because the Li-Ge system suffers from large volume changes during the Li insertion and removal process, which can cause cracking of the electrode and a loss of electrical contact. The use of nanostructures instead of bulk Ge structures decreases the mechanical strains on the system, which can lead to increased capacity and cycle life.

In battery studies, both crystalline and amorphous phases have been ob-

served during cycling [60, 27]. The crystal structures observed while charging and discharging Li-Ge batteries vary between different studies. In particular, the identity of the fully-lithiated phase has been debated. The crystalline $Li_{4.4}Ge$ phase is often used to estimate the system's theoretical capacity, and the presence of a phase with this composition has been reported [60]. However, this structure has also been identified as $Li_{21.1875}Ge_5$ [59]. Oudenhoven *et al.* were unable to experimentally reach a higher lithium content than 3.75 Li per Ge, which corresponds to the $Li_{15}Ge_4$ phase [111]. Available phase diagrams disagree further as to the ground state binary phases at low temperatures [109, 132, 20]. In summary, the phases LiGe, $Li_{11}Ge_6$, Li_9Ge_4 , Li_7Ge_2, $Li_{15}Ge_4$, Li_7Ge_{12}, and $Li_{21.1875}Ge_5$ have experimentally been observed. The last two, Li_7Ge_{12} and $Li_{21.1875}Ge_5$, are reportedly to exhibit partial site occupancies, and in this work we study several fully occupied versions that may form at 0 K.

To better understand the Li-Ge materials system, we calculate the energies of the various known crystalline Li-Ge structures using density-functional theory. This allows us to predict the voltage characteristics and charge capacity of a germanium anode. Furthermore, we search the phase diagram using the Genetic Algorithm for Structure and Phase Prediction (GASP) code [151, 130, 150]. This search identifies a structure with composition Li_5Ge_2 which does not appear in previous phase diagrams but which we predict to be thermodynamically stable with respect to all competing phases at low temperature. We characterize this new material and discuss its relationship to other binary Li-Ge materials.

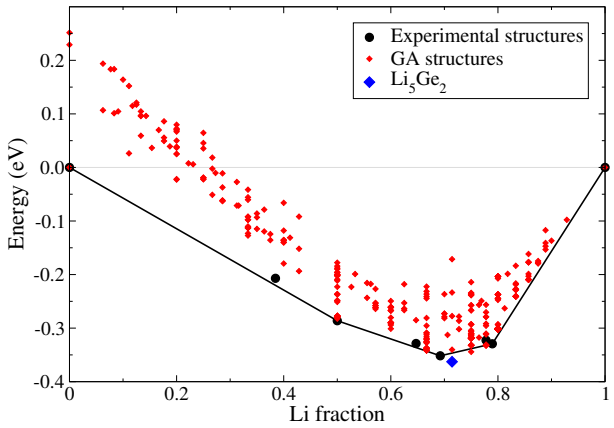


Figure 4.6: (Color online.) Formation energies as a function of Li content for Li-Ge structures. The Li_5Ge_2 structure identified by the genetic algorithm search is stable relative to the known competing experimental structures.

4.4 Methods

Density-functional theory (DFT) calculations are performed using the VASP package [85, 84], a plane-wave code using the projector-augmented wave method [18, 86] and the Perdew-Burke-Ernzerhof (PBE) generalized-gradient approximation to the exchange-correlation functional [117]. For Ge, the 4*s* and 4*p* states are treated as valence states and for Li, all three electrons are treated as valence states. A plane-wave energy cutoff of 350 eV and a *k*-point mesh with a density of 40 per $^{-1}$ ensure energy convergence to within 1 meV/atom. The Brillouin-zone integration is performed with the Methfessel-Paxton scheme as implemented in VASP with a smearing of 0.1 eV.

An *ab-initio* structure search was performed using the GASP code [151, 130, 150] to explore the low-energy structures of the Li-Ge phase diagram. It uses a grand-canonical genetic algorithm to find structures, which minimize the formation energy with respect to elemental Li and Ge for all compositions between

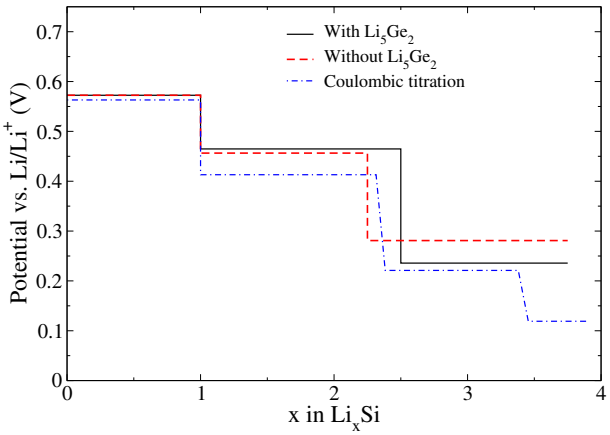


Figure 4.7: (Color online.) Potential of the Li-Ge system as a function of Li content relative to the potential of Li/Li^+. The Coulombic titration data from Ref. [131] was taken at 420 °C.

those two endpoints. It works by creating successive generations of candidate structures. Each candidate is relaxed and its total energy calculated using DFT as described above. High-energy structures are discarded, while low-energy ones are maintained or used to create more candidates. In this way, properties of the candidate structures which lead to low energy are propagated in the population, while unfavorable features die out. About 500 total energy calculations and structural relaxations were performed in this search. For more information about the method, see Ref. [151].

The Li-Ge binary phase diagram is created from the computed formation energies using the convex hull construction [151]. The voltage curve describing the potential of a Li/Ge anode as a function of lithium content with respect to Li/Li^+ was computed from the energies of the ground state phases [10, 148]. The method of finite displacements as implemented in the Phonopy package [153] interfaced with VASP was used to calculate the phonon spectrum and verify the dynamical stability of the Li_5Ge_2 structure. The elastic constants and electronic density of states of the Li_5Ge_2 phase were also computed with VASP.

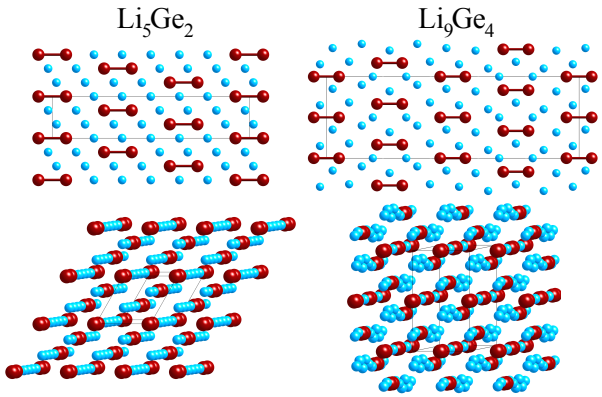


Figure 4.8: The Li_5Ge_2 and Li_9Ge_4 compounds are structurally similar.

Powder diffraction patterns were computed using the CRYSTALDIFFRACT software package [1].

4.5 Results

4.5.1 Stability and electrical potential of Li-Ge phases

Figure 4.6 shows the formation energies of the experimental structures as well as those encountered by the genetic algorithm structure search. The formation energies of these structures are mostly above the convex hull of the experimental structures and hence metastable. However, the genetic algorithm identifies a structure with composition Li_5Si_2 that is predicted to be stable with respect to the experimentally-observed phases. LiGe and $Li_{15}Ge_4$ are thermodynamically stable. The Li_9Ge_4 phase is predicted to be stable with respect to the other experimentally known structures but is destabilized by the new Li_5Ge_2 phase. Li_7Ge_{12}, $Li_{11}Ge_6$, and Li_7Ge_2 lie close to the convex hull, all within 15 meV/atom of the ground states. Thus, these structures may be entropically stabilized at

higher temperatures. None of our fully occupied versions of the partially occupied $Li_{22}Ge_5$ phase lay near the convex hull. This indicates that $Li_{22}Ge_5$ phase is an entropically stabilized high-temperature phase. **[RGH: After updating the convex hull plot, change this sentence if needed.]**

Fig. 4.7 compares the calculated adiabatic electric potential of the system over the course of lithiation/delithiation with and without the Li_5Si_2 phase with the experimental data reported in the patent application by Sammells *et al.* for a Coulombic titration experiment performed at 420 °C [131]. Our predictions compare favorably to the data, despite the fact that the Li-Ge structures observed during the loading and unloading of Ge in electrochemical cells varies widely, as discussed above. In fact, a phase with composition Li_5Ge_2 is noted in the Coulombic titration data. However, its structure is not identified, and it is not noted in any of the later phase diagrams cited above.

4.5.2 The Li_5Ge_2 phase

Table 4.2 lists the structural parameters of the new Li_5Ge_2 crystal structure, optimized with DFT. Fig. 4.8 illustrates the Li_5Ge_2 structure and compares it to the Li_9Ge_4 structure. Both structures are dominated by chains of Li atoms that contain Ge dimers indicating a Zintl-like stabilization mechanism [104]. The Li_5Ge_2 structure consists of straight parallel chains of 5 Li atoms and 2 Ge dimers. Similarly, the Li_9Ge_4 structure is composed of chain-like structural motives with Ge dimers. However, the chains in Li_9Ge_4 are not straight and contain a larger number Li atoms.

Fig. 4.9 compares the predicted powder diffraction patterns of both struc-

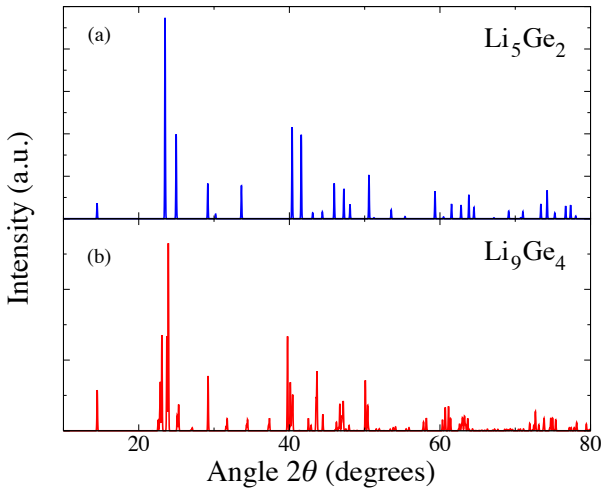


Figure 4.9: Comparison of the calculated powder diffraction patterns of the predicted Li_5Ge_2 structure with the nearby experimentally observed phase Li_9Ge_4 phase with a similar crystal structure.

tures. Although their geometries are qualitatively similar, the diffraction patterns appear to differ sufficiently as to allow experiments to distinguish both phases even for samples with small grain sizes and disorder. It seems therefore unlikely that the structure was misidentified in the experimental studies. Unfortunately, the experimental diffraction pattern was not reported in Ref. [70]. Our DFT energies predict the Li_9Ge_4 structure to be slightly unstable at low temperatures against decomposition into the Li_5Ge_2 and the LiGe structure as shown in Fig. 4.6. However, the energy of the Li_9Ge_4 structure is only a few meV above the 0 K convex and thus may be entropically stabilized at higher temperatures.

To determine the bonding characteristics of the Li_5Ge_2 compound, we calculate the electronic structure of the compound. The Bader charge analysis shows that the Ge ions have a charge of -2 and the Li ions a charge of $+0.8$. This is consistent with a Zintl compound composed of Li^+ cations and $(Ge_2)^{-4}$ Zintl anions where the remaining additional electron is delocalized on the Li sites. Figure 4.10 shows the electronic band structure of the Li_5Ge_2 compound projected

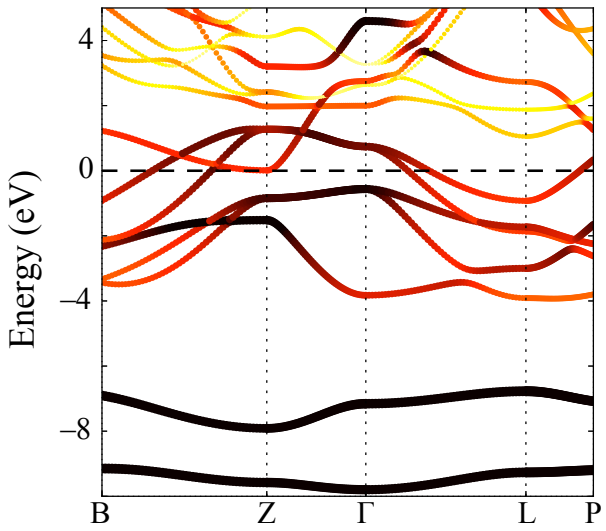


Figure 4.10: Electronic band structure of the Li_5Ge_2 compound projected on the Ge states.

on the Ge sites. The Li_5Ge_2 compound is metallic with several bands crossing the Fermi level. The states at the Fermi level are dominated by the Ge *p* states and the Li *s* states are largely shifted to higher energy. The Fermi level falls within a broad pseudogap region with a reduced density of states, consistent with the Zintl-type stabilization mechanism.

When Ge is considered for application as an anode in Li-ion batteries, a large

Table 4.2: Structure of Li_5Ge_2. Its space group is $R\bar{3}m$ (166) and its lattice parameters are a = 4.464 and c = 18.353 .

Site	x	y	z
Li1 1(b)	0.0	0.0	0.5
Li2 2(c)	0.0	0.0	0.354
Li3 2(c)	0.0	0.0	0.212
Ge 2(c)	0.0	0.0	0.070

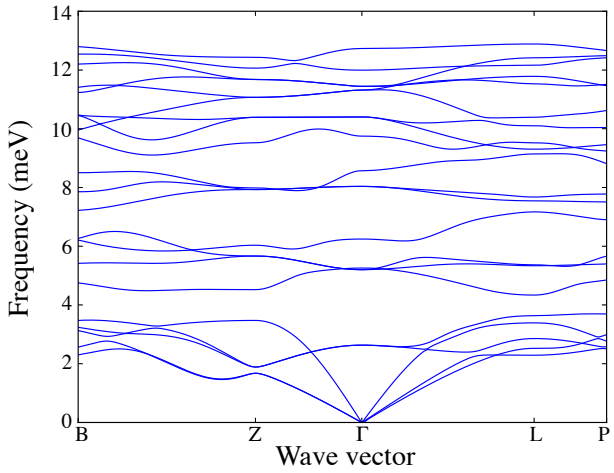


Figure 4.11: The phonon spectrum of the Li_5Ge_2 compound.

volume change has to be accommodated during the charge/discharge cycle. Hence it is important to know the elastic properties of the various phases that can occur. Table 4.3 provides the elastic constant tensor of the Li_5Ge_2 compound. We find that the elastic constants of Li_5Ge_2 are significantly larger the ones of pure bcc Li [36] and similar to pure Ge [19]. This is consistent with the above observation that despite the high Li content, the bonding in Li_5Ge_2 differs from the metallic bonding of pure Li. Fig. 4.11 shows the phonon spectrum of the Li_5Ge_2 structure along the reciprocal space path as suggested by Setyawan *et al.* [137]. The absence of imaginary modes demonstrates that Li_5Ge_2 structure is dynamically stable. The phonon frequencies of Li_5Ge_2 are larger than for bcc Li, which is consistent with the higher elastic constants and a Zintl-type compound.

4.6 Conclusions

The theoretical capacity of the Li-Ge system is much higher than that of currently used graphitic anode materials. We have performed calculations to iden-

tify the stable low-temperature crystal structures of the Li-Ge system. Our predicted phase diagram is consistent with previously reported experimental studies. A voltage profile for the system indicates that a germanium anode will have a reasonably low potential for use as a lithium-ion battery anode.

Additionally, we identify a binary with composition Li_5Ge_2 not included in previous phase diagrams, which we predict will be a thermodynamic ground state at low temperature. A crystal structure with this composition was noted in previous Coulombic titration data but was not further characterized. In this paper, we present structural data, a phonon spectrum, the electronic density of states, the powder diffraction pattern, and the tensor of elastic constants for this material.

Table 4.3: The elastic constant tensor c_{ij} in GPa for the Li_5Ge_2 structure.

	xx	yy	zz	xy	yz	zx
xx	95	16	-8	0	-140	0
yy		95	-8	0	140	0
zz			133	0	0	0
xy				39	0	-140
yz					18	0
zx						18

CHAPTER 5

TESTING EMPIRICAL POTENTIALS WITH STOCHASTIC SEARCH

Much of the content of this chapter was previously published in Refs. [151] and [114].

Many modern potentials, such as spline-based MEAM models, employ highly-flexible functional forms with numerous parameters in order to closely fit training data. However, this flexibility increases the risk of stabilizing unexpected structures. It is very challenging to avoid unphysical minima or low-energy local minima when designing potentials without explicitly checking for them with a method capable of identifying unforeseen structures. Additionally, performing such checks requires a complete search over all compositional degrees of freedom. Simply verifying that a potential's lowest-energy structures at compositions known to hold ground states is insufficient, since they can easily be destabilized by false low-lying minima at other compositions.

We have found that searching the energy landscape induced by an empirical potential is an important step in verifying the potential itself. The presence of incorrect ground states or low-energy local minima is indicative of unphysical interactions which may be problematic when the potential is used in applications such as molecular dynamics (MD) simulations. In this chapter, we describe tests of two empirical potential energy models.

5.1 Molybdenum

Ref. [114] describes the construction and testing of an empirical energy model describing elemental molybdenum. The GASP code was used to search for low-

Table 5.1: This table from Ref. [114] shows the MEAM values for the cohesive energy, lattice parameter, bulk modulus, and elastic constants of bcc Mo are compared to DFT and experiment. The energies and lattice parameters of the fcc, hcp, β-W, β-Ta, and ω-Ti structures are compared to DFT results. The energies are relative to the energy of the bcc structure.

	MEAM[1]	GGA-PBE[1]	Experiment
E_{coh} (eV/atom)	6.82	6.25	6.82[2]
a ()	3.167	3.169	3.147[3]
B (GPa)	253	263	270[4]
C_{11} (GPa)	441	462	479[4]
C_{12} (GPa)	158	163	165[4]
C_{44} (GPa)	96	102	108[4]
$\Delta E_{\text{monoclinic-bcc}}$ (meV/atom)	198	175	...
$\Delta E_{\text{Fddd-bcc}}$ (meV/atom)	242	231	...
$\Delta E_{\beta\text{W-bcc}}$ (meV/atom)	266	96	...
$a_{\beta\text{W}}$ ()	5.026	5.058	...
$\Delta E_{\text{Pmma-bcc}}$ (meV/atom)	269	233	...
$\Delta E_{\beta\text{Ta-bcc}}$ (meV/atom)	280	168	...
$a_{\beta\text{Ta}}$ ()	9.719	9.752	...
$c_{\beta\text{Ta}}$ ()	5.048	5.113	...
$\Delta E_{\omega\text{Ti-bcc}}$ (meV/atom)	332	404	...
$a_{\omega\text{Ti}}$ ()	4.616	4.681	...
$c_{\omega\text{Ti}}$ ()	2.595	2.572	...
$\Delta E_{\text{fcc-bcc}}$ (meV/atom)	391	418	...
a_{fcc} ()	3.931	4.013	...
$\Delta E_{\text{hcp-bcc}}$ (meV/atom)	415	433	...
a_{hcp} ()	2.743	2.765	...
c_{hcp} ()	4.692	4.905	...

energy structures with respect to this MEAM potential. This search this allows us to check that our potential reproduces the known true ground state (a BCC crystal) and to identify low energy metastable phases. Unit cells of up to 40 atoms were considered, and the search ran for 100 generations with 100 distinct candidate structures in each generation. The BCC and various defect and metastable structures were encountered by the algorithm many times. The BCC ground state was indeed the lowest energy structure found.

In addition to many structures representing defects in a BCC cell, several notable metastable crystals were identified by the genetic algorithm. The lowest energy of these is a 5-atom monoclinic structure. Its formation energy per atom is +198 meV compared to BCC according to our potential and +175 meV according to DFT. Next, we find a 4-atom structure with space group 51 (Pmma) and with energy of +242 meV accroding to the potential and +231 meV according to DFT with respect to the ground state. Finally, the A15 (beta-W) phase is +266 meV or +96 meV above BCC, and a 2-atom structure with space group 70 (Fddd) has energy +269 meV or +233 meV above the groudn state according to our potential and DFT, respectively. Table 5.1 includes a summary of these structures.

We found no other crystal structures within 270 meV of the ground state. This result provides very strong indication that this potential has the correct ground state and that no other crystal phase should occur in any MD simulations for pressure and temperature ranges as described in this chapter. The potential's reproduction of the correct ground state and accurate description of the low-lying metastable structures indicate that the potential captures many of the important properties of the *ab-initio* potential energy landscape.

5.2 Cu-Zr-Al

5.2.1 Methodology

To test the performance of the GA for phase diagram predictions and to determine the accuracy of empirical potentials, we apply the GA to the Zr-Cu-Al ternary system for which an EAM empirical potential is available.[30] Testing the EAM potential requires the knowledge of the true ground state structures of the ternary Zr-Cu-Al system. This phase diagram is reviewed in Ref. [128]. Structural data for each ternary, binary, and elemental phase was extracted from the Inorganic Crystal Structure Database[15] (ICSD) whenever possible and from Refs. [2] and [42]. The phase diagram indicates 10 ternary structures labelled τ_1 through τ_{10}. Unfortunately, refined structural data appears to be unavailable for $\tau_1 - \tau_3$, τ_6, τ_8, and τ_9. These may be entropically-stabilized phases which do not tend to form at low temperatures. The structures τ_5 and τ_7 are reported with partial occupancy, likely a result of the high temperature conditions at which these phases were synthesized. We generated several fully-occupied candidate structures based on the reported structures for τ_5 and τ_7.

Density functional theory (DFT) calculations are used to identify our best estimate of the ground state energetics of the Zr-Cu-Al ternary system. That is, the structures from the literature as well as all of the EAM ground states identified by the GA are evaluated using DFT and a convex hull is calculated. The DFT calculations are performed with VASP [85, 84], a density functional code which uses a plane-wave basis and the projector-augmented wave method.[18, 86] The k-point density and plane wave cutoff energy are chosen to ensure convergence of the energies to within 2 meV/atom. We will refer to the structures on this

DFT convex hull as the experimental ground states.

We similarly predict a set of EAM ground states. A single long run of 100 generations which encountered about 8000 unique structures was used to explore the EAM energy landscape. This run was independent of those used to generate the performance distributions, but we verified that no structures from the shorter performance distribution runs would improve this run's predicted convex hull. The experimentally-known structures and the set of structures from this GA search are evaluated using the EAM potential and a convex hull constructed. We will refer to the phases on this convex hull as the EAM ground states.

In order to evaluate the accuracy of the EAM potential, we will look at two sets of energies of formation with respect to the ground states. First we have the EAM formation energies of the experimental ground states. Geometrically, these are the distances above the EAM convex hull that the experimental ground states lie, according to the EAM energy model. Equivalently, we can think of these as how far below the correct experimental structures the potential predicts some unphysical structures to lie. The larger these energies, the worse the behavior of the empirical potential. The fact that we have no proof that our sets of structures are ground states of their respective Hamiltonians is not really problematic. The DFT ground states are experimentally-verified, so they should be ground states of the true Hamiltonian, and any further searching of the EAM or DFT energy landscapes could only result in lower-energy EAM structures found which will only make these erroneous formation energies even larger.

Second, we find the DFT formation energies of the EAM ground states with respect to the DFT convex hull. The larger these energies, the worse the struc-

tures stabilized by the potential are according to our benchmark calculations. In the ideal case that the potential accurately describes the experimental ground states of the system, all of these energy differences would be zero.

In both sets of calculations (i.e. when evaluating DFT structures with EAM and evaluating EAM structures with DFT) we fully re-optimize the geometries of each structure using the latter energy model, including both the lattice and atomic degrees of freedom. This provides structures with the opportunity to lower their energies by relax to potentially quite different structures. Thus, this method of comparison could potentially mask particularly egregious errors as when the potential stabilizes structures which are extremely unstable according to the *ab-initio* calculations. So in this sense, our methodology is somewhat generous to the empirical potential.

5.2.2 Results

The calculation of the energy landscape using a genetic algorithm provides a stringent test for empirical potentials[114] that can reveal numerical instabilities, unphysical structures, and incorrect local or global minima structures. Initial genetic algorithm searches discovered that the Zr-Cu-Al EAM potential stabilizes many structures with unphysically short interatomic distances and diverges when applied to others.

We rectified this issue by explicitly excluding any structures with interatomic distances less than 1.5 . Such hard constraints are difficult to apply in applications such as MD simulations and these results are indicative of unphysical interactions in certain regimes. It could be the case that energy barriers exist

which keep a system from arriving in these configurations in an MD simulation assuming the calculation starts in a physical state. However, this defect is one that, if recognized during the design phase, can easily be fixed in an EAM potential by modification of the pair potential term,[65] and such issues would be discovered if a structure search was used as part of potential design and testing methodology.[114]

The potential's elemental ground states for Zr and Cu are hcp and fcc, respectively, and agree with experiment. However, the true ground state of Al is fcc whereas the EAM potential predicts the hcp structure lies 5 meV/atom lower in energy. (NB: We verified this issue with the authors of the potential.) In this case, we might expect the potential to predict incorrect mechanical properties for Al-rich systems since the stacking fault energies are incorrect.

Table 5.2 describes the EAM ground states. It gives their DFT formation energies with respect to the DFT ground states as well as their EAM formation energies with respect to the elements. All of these structures (except fcc Al) are on the convex hull of the EAM phase diagram, and any with DFT formation energies, E_{form}, greater than zero are not on the convex hull of the DFT phase diagram. We observe that majority of the ground states predicted by the EAM potential are unstable in DFT relative to the experimental phases. Note again, that even if the phases discovered by the GA with the EAM potential are not the true EAM ground states, the energies of these phases are below the energy of the experimental phases on the EAM energy landscape. It is encouraging to see that none of the formation energies are particularly large, indicating that the overall energy order of the various phases is similar in DFT and EAM. This is, however, expected since the structures are all re-relaxed using DFT. These

relaxations can be significant and for several of the EAM-stabilized geometries, the DFT relaxations even change the symmetries of the structures.

Table 5.3 similarly describes the DFT ground states. It gives their EAM formation energies with respect to the EAM convex hull as well as their DFT formation energies with respect to the elements. When these are non-zero, it indicates that low-energy structures at the respective composition will tend to form the wrong geometries and thus often display incorrect properties as in the case of elemental Al. We find that many of the experimental binary phases are not stable in the EAM potential. A re-optimization of the EAM potential using this data could overcome some of these issues. The inaccuracies uncovered by the genetic algorithm search illustrate that the development of empirical potentials for multi-component systems is an ambitious, difficult, and still unsolved problem.

Table 5.2: EAM ground states: DFT formation energies with respect to the DFT ground states, $E_t^{\mathrm{gs}} extrm form$, and EAM formation energies with respect to the elements, E_{form} (meV/atom).

Composition	Space group		$E_{\mathrm{form}}^{\mathrm{gs}}$	E_{form}
Al (hcp)	194	$P6_3/mmc$	27	0
Al (fcc)	225	$Fm\bar{3}m$	0	5
Cu	225	$Fm\bar{3}m$	0	0
Zr	194	$P6_3/mmc$	0	0
ZrCu	221	$Pm\bar{3}m$	51	-155
ZrAl	65	$Cmmm$	57	-542
$ZrAl_2$	191	$P6/mmm$	101	-517
Zr_2Al	194	$P6_3/mmc$	25	-403
Zr_3Al	59	$Pmmn$	45	-315
$ZrAl_6$	12	$C2/m$	48	-300
$ZrAl_8$	12	$C2/m$	77	-239
Zr_2Cu	139	$I4/mmm$	0	-115
Zr_3Cu_8	62	$Pnma$	0	-118
CuAl	221	$Pm\bar{3}m$	70	-330
$CuAl_2$	65	$Cmmm$	53	-287
Cu_3Al	139	$I4/mmm$	3	-187
Zr_3Al_2	166	$R\bar{3}m$	129	-463
$ZrAl_3$	221	$Pm\bar{3}m$	26	-464
$ZrAl_5$	191	$P6/mmm$	118	-344
$CuAl_4$	87	$I4/m$	87	-181
$ZrCuAl_2$	123	$P4/mmm$	95	-445
Zr_2Cu_2Al	71	$Immm$	54	-325
$Zr_2Cu_2Al_3$	12	$C2/m$	36	-446
$Al_7Cu_{16}Zr_6$	225	$Fm\bar{3}m$	0	-335

Table 5.3: DFT ground states: EAM formation energies with respect to the EAM ground states, $E_t^{\mathrm{gs}} extrmform$, and DFT formation energies with respect to the elements, E_{form} (meV/atom).

Composition	Space group		$E_t^{\mathrm{gs}} extrmform$	E_{form}
Zr	194	$P6_3/mmc$	0	0
Al (fcc)	225	$Fm\bar{3}m$	5	0
Cu	225	$Fm\bar{3}m$	0	0
$ZrAl_2$	194	$P6_3/mmc$	28	-543
Zr_4Al_3	191	$P6/mmm$	89	-479
Zr_2Cu	139	$I4/mmm$	0	-141
Zr_3Cu_8	62	$Pnma$	0	-169
Zr_7Cu_{10}	41	$Aba2$	33	-166
$Zr_{14}Cu_{51}$	174	$P\bar{6}$	15	-164
Al_3Zr	139	$I4/mmm$	15	-487
$AlZr_3$	221	$Pm\bar{3}m$	5	-304
AlCu	12	$C2/m$	87	-218
Al_2Cu	140	$I4/mcm$	44	-162
$AlCu_3$	59	$Pmmn$	4	-190
Al_4Cu_9	215	$P\bar{4}3m$	21	-216
$AlCu_2Zr$	225	$Fm\bar{3}m$	26	-359
$ZrCu_6Al_6$	6	Pm	61	-322
$Al_7Cu_{16}Zr_6$	225	$Fm\bar{3}m$	0	-351

CHAPTER 6

CONCLUSIONS

We have described our approach to structure and phase diagram prediction using a genetic algorithm. Our code, the Genetic Algorithm for Structure and Phase Prediction (GASP), is freely available and also supports searching for molecules and atomic clusters. It includes tools for structure visualization, construction and manipulation of convex hulls, and generation of voltage profiles.

We have described three classes of applications. First, we considered the elemental europium and barium under pressure. These studies were motivated by the need to better understand the origins of the superconducting transitions in these materials. Each of these studies provides a good example of a situation where a computational approach compliments experiment and represents a textbook applications of our search methodology: in each case the goal was to determine a crystal structure under extreme conditions where experiment id difficult.

However, we have seen that structure determination is also an important step in more involved computational materials studies. We presented a methodology for screening materials for use as Li-ion battery electrodes, and we focused on the Li-Si and Li-Ge battery electrode systems as examples. The primary challenge in performing such characterization from first principles is in determining the structure of the material during charging and discharging of the battery.

Thirdly, we have discussed here the use of the search method to test empirical potentials. We considered a widely-used EAM potential for the Zr-Cu-Al

system and found evidence of non-physical interactions which could have been identified during the potentials' design phase using an explicit search of the energy landscape such as we have described here. We considered the DFT energies of the EAM ground states and the EAM energies of the DFT ground states. Neither of these studies is possible without an explicit structure search method, and we and identified a number of errors on the order of 100 meV/atom in both cases.

Several avenues of methodological improvement are ongoing. First, prediction of surfaces and 2-D materials is in progress. Second, we explore the use of surrogate energy models to reduce the number of *ab-initio* calculations necessary. These surrogate energy models may be either pre-existing empirical or semi-empirical models or new models dynamically generated over the course of a GA run. This approach should be contrasted with another tack on reducing energy calculations, smart variation operators[95], and also with methods which perform their entire search process on an inexpensive model and only evaluate their results with DFT.

BIBLIOGRAPHY

[1] Images and video generated using CrystalMaker: a crystal and molecular structures program for Mac and Windows. CrystalMaker Software Ltd, Oxford, England (www.crystalmaker.com).

[2] Al-cu-zr (aluminium - copper - zirconium). In G. Effenberg and S. Ilyenko, editors, *Light Metal Systems. Part 2*, volume 11A2 of *Landolt-Börnstein - Group IV Physical Chemistry*, pages 1–17. Springer Berlin Heidelberg, 2005.

[3] N. L. Abraham and M. I. J. Probert. A periodic genetic algorithm with real-space representation for crystal structure and polymorph prediction. *Phys. Rev. B*, 73:224104, Jun 2006.

[4] N. L. Abraham and M. I. J. Probert. Improved real-space genetic algorithm for crystal structure and polymorph prediction. *Phys. Rev. B*, 77:134117, Apr 2008.

[5] Maximilian Amsler and Stefan Goedecker. Crystal structure prediction using the minima hopping method. *J. Chem. Phys.*, 133(22):224104, 2010.

[6] Qi An, Sheng-Nian Luo, William A. Goddard, III, W. Z. Han, B. Arman, and William L. Johnson. Synthesis of single-component metallic glasses by thermal spray of nanodroplets on amorphous substrates. *Appl. Phys. Lett.*, 100(4):041909, 2012.

[7] R. L. Anderson. Recent advances in finding best operating conditions. *Journal of the American Statistical Association*, 48, 1953.

[8] R. S. Anderssen and P. Bloomfield. Properties of the random search in global optimization. *Journal of Optimziation Theory and Applications*, 16, 2005.

[9] B. Arman, C. Brandl, S. N. Luo, T. C. Germann, A. Misra, and T. Cagin. Plasticity in cu(111)[bold /]cu[sub 46]zr[sub 54] glass nanolaminates under uniaxial compression. *J. Appl. Phys.*, 110(4):043539, 2011.

[10] M. K. Aydinol, A. F. Kohan, G. Ceder, K. Cho, and J. Joannopoulos. *Ab initio* study of lithium intercalation in metal oxides and metal dichalcogenides. *Phys. Rev. B*, 56:1354–1365, Jul 1997.

[11] M.S. Bailey, N.T. Wilson, C. Roberts, and R.L. Johnston. Structures, stabilities and ordering in ni-al nanoalloy clusters. *European Physical Journal D*, 25, 2003.

[12] S. Baroni, A. Dal Corso, S. de Gironcoli, P. Giannozzi, C. Cavazzoni, G. Ballabio, S. Scandolo, G. Chiarotti, P. Focher, A. Pasquarello, K. Laasonen, A. Trave, R. Car, N. Marzari, and A. Kokalj. Pwscf. `http://www.pwscf.org/`.

[13] I. Barvik. To the magnetic properties of Li_2.33Si. *Czech. J. Phys. B*, 33:1338, 1983.

[14] Enrique R. Batista, Jochen Heyd, Richard G. Hennig, Blas P. Uberuaga, Richard L. Martin, Gustavo E. Scuseria, C. J. Umrigar, and John W. Wilkins. Comparison of screened hybrid density functional theory to diffusion monte carlo in calculations of total energies of silicon phases and defects. *Phys. Rev. B*, 74:121102, Sep 2006.

[15] Alec Belsky, Mariette Hellenbrandt, Vicky Lynn Karen, and Peter Luksch. New developments in the Inorganic Crystal Structure Database (ICSD): accessibility in support of materials research and design. *Acta Cryst. B*, 58(3 Part 1):364–369, Jun 2002.

[16] R. Stephen Berry. Potential surfaces and dynamics: What clusters tell us. *Chemical Review*, 93:2379–2394, 1993.

[17] W. Bi, Y. Meng, R. S. Kumar, A. L. Cornelius, W. W. Tipton, R. G. Hennig, Y. Zhang, C. Chen, and J. S. Schilling. Pressure-induced structural transitions in europium to 92 gpa. *Phys. Rev. B*, 83:104106, Mar 2011.

[18] P. E. Blöchl. Projector augmented-wave method. *Phys. Rev. B*, 50:17953–17979, Dec 1994.

[19] W. L. Bond, W. P. Mason, H. J. McSkimin, K. M. Olsen, and G. K. Teal. The elastic constants of germanium single crystals. *Phys. Rev.*, 78:176–176, Apr 1950.

[20] Hans Ulrich Borgstedt, Cezary Guminski, Hans Ulrich Borgstedt, and Cezary Guminski. Iupac-nist solubility data series. 75. nonmetals in liquid alkali metals. *Journal of Physical and Chemical Reference Data*, 30(4):835–1158, 2001.

[21] L. Brewer. The cohesive energies of the elements. *Lawrence Berkeley Laboratory Report*, LBL-3720, May 1977.

[22] T. S. Bush, C. R. A. Catlow, and P. D. Battle. Evolutionary programming techniques for predicting inorganic crystal structures. *J. Mater. Chem.*, 5, 1995.

[23] T. S. Bush, C. R. A. Catlow, and P. D. Battle. Evolutionary programming techniques for predicting inorganic crystal structures. *J. Mater. Chem.*, 5:1269–1272, 1995.

[24] A.J. Cao, Y.Q. Cheng, and E. Ma. Structural processes that initiate shear localization in metallic glass. *Acta Mater.*, 57(17):5146 – 5155, 2009.

[25] Gerbrand Ceder, Dane Morgan, Chris Fischer, Kevin Tibbetts, and Stefano Curarolo. Data-mining-driven quantum mechanics for the prediction of structure. *MRS Bulletin*, 31, 2006.

[26] Candace K. Chan, Hailin Peng, Gao Liu, Kevin McIlwrath, Xiao Feng Zhang, Robert A. Huggins, and Yi Cui. High-performance lithium battery anodes using silicon nanowires. *Nature Nanotechnology*, 3:31–35, 2008.

[27] Candace K. Chan, Xiao Feng Zhang, and Yi Cui. High capacity li ion battery anodes using ge nanowires. *Nano Letters*, 8(1):307–309, 2008.

[28] Y. Q. Cheng, J. Ding, and E. Ma. Local topology vs. atomic-level stresses as a measure of disorder: Correlating structural indicators for metallic glasses. *Materials Research Letters*, 1:3–12, 2013.

[29] Y. Q. Cheng and E. Ma. Configurational dependence of elastic modulus of metallic glass. *Phys. Rev. B*, 80:064104, Aug 2009.

[30] Y. Q. Cheng, E. Ma, and H. W. Sheng. Atomic level structure in multicomponent bulk metallic glass. *Phys. Rev. Lett.*, 102:245501, Jun 2009.

[31] V. L. Chevrier and J. R. Dahn. First principles model of amorphous silicon lithiation. *Journal of The Electrochemical Society*, 156(6):A454–A458, 2009.

[32] V. L. Chevrier and J. R. Dahn. First principles studies of disordered lithiated silicon. *Journal of The Electrochemical Society*, 157(4):A392–A398, 2010.

[33] V. L. Chevrier, J. W. Zwanziger, and J. R. Dahn. First principles studies of silicon as a negative electrode material for lithium-ion batteries. *Can. J. Phys.*, 87(6):625–632, 2009.

[34] V.L. Chevrier, J.W. Zwanziger, and J.R. Dahn. First principles study of lisi crystalline phases: Charge transfer, electronic structure, and lattice vibrations. *Journal of Alloys and Compounds*, 496(12):25 – 36, 2010.

[35] Gabor Csanyi, Chris J Pickard, B D Simons, and R J Needs. Graphite intercalation compounds under pressure: A first-principles density functional theory study. *Physical Review B*, 75, 2007.

[36] J. P. Day and A. L. Ruoff. The variation of the elastic constants of lithium with temperature and pressure. *physica status solidi (a)*, 25(1):205–213, 1974.

[37] D. M. Deaven and K. M. Ho. Molecular geometry optimization with a genetic algorithm. *Phys. Rev. Lett.*, 75:288–291, Jul 1995.

[38] M. Debessai, T. Matsuoka, J. J. Hamlin, J. S. Schilling, and K. Shimizu. Pressure-induced superconducting state of europium metal at low temperatures. *Phys. Rev. Lett.*, 102:197002, May 2009.

[39] J. Ding, Y. Q. Cheng, and E. Ma. Correlating local structure with inhomogeneous elastic deformation in a metallic glass. *Appl. Phys. Lett.*, 101(12):121917, 2012.

[40] Jun Ding, Yong-Qiang Cheng, Hongwei Sheng, and Evan Ma. Short-range structural signature of excess specific heat and fragility of metallic-glass-forming supercooled liquids. *Phys. Rev. B*, 85:060201, Feb 2012.

[41] T. Egami. Mechanical failure and glass transition in metallic glasses. *J. Alloys Compd.*, 509(0):S82 – S86, 2011.

[42] Mohamed El-Boragy, Reiner Szepan, and Konrad Schubert. Kristallstruktur von cu3al2+ (h) und cual (r). *J. Less Common Met.*, 29(2):133 – 140, 1972.

[43] A. Gauzzi et. al. Maximum tc at the verge of a simultaneous order-disorder and lattice-softening transition in superconducting cac6. *Physical Review B*, 78, 2008.

[44] I. Chaudhuri et. al. Global optimization of silicon nanoclusters. *Applied Surface Science*, 226, 2004.

[45] J P M Lommerse et. al. A test of crystal structure prediction of small organic molecules. *Acta Cryst.*, B56, 2000.

[46] W D S Motherwell et. al. Crystal structure predition of small organic molecules: a second blind test. *Acta Cryst.*, B58, 2002.

[47] William H. Press et. al. *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, 3 edition, 2007.

[48] Vinodkumar Etacheri, Rotem Marom, Ran Elazari, Gregory Salitra, and Doron Aurbach. Challenges in the development of advanced li-ion batteries: a review. *Energy Environ. Sci.*, 4:3243–3262, 2011.

[49] Alessandro Fadda and Giuseppe Fadda. An evolutionary algorithm for the prediction of crystal structures. *Phys. Rev. B*, 82:104105, Sep 2010.

[50] Ji Feng, Richard G Hennig, N W Ashcroft, and Roald Hoffman. Emergent reduction of electronic state dimensionality in dense ordered li-be alloys. *Nature*, 451, 2008.

[51] Christopher C. Fischer, Kevin J. Tibbetts, Dane Morgan, and Gerbrand Ceder. Predicting crystal structure by merging data mining with quantum mechanics. *Nat. Mater.*, 5:641–646, 2006.

[52] D. J. Fredeman, P. H. Tobash, M. A. Torrez, J. D. Thompson, E. D. Bauer, F. Ronning, W. W. Tipton, Sven P. Rudin, and R. G. Hennig. Computationally driven experimental discovery of the ceir$_4$in compound. *Phys. Rev. B*, 83:224102, Jun 2011.

[53] Eric Freeman, Elisabeth Robson, Bert Bates, and Kathy Sierra. *Head First Design Patterns*. O'Reilly Media, 2004.

[54] T. Fujita, P. F. Guan, H. W. Sheng, A. Inoue, T. Sakurai, and M. W. Chen. Coupling between chemical and dynamic heterogeneities in a multicomponent bulk metallic glass. *Phys. Rev. B*, 81:140204, Apr 2010.

[55] Takeshi Fujita, Zheng Wang, Yanhui Liu, Howard Sheng, Weihua Wang, and Mingwei Chen. Low temperature uniform plastic deformation of

metallic glasses during elastic iteration. *Acta Mater.*, 60(9):3741 – 3747, 2012.

[56] Julian D Gale and Andrew L Rohl. The general utility lattice program (gulp). *Molecular Simulation*, 29, 2003.

[57] Colin W. Glass, Artem R. Oganov, and Nikolaus Hansen. Uspexevolutionary crystal structure prediction. *Computer Physics Communications*, 175(1112):713 – 720, 2006.

[58] Aurlien Gohier, Barbara Laik, Jean-Pierre Pereira-Ramos, Costel Sorin Cojocaru, and Pierre Tran-Van. Influence of the diameter distribution on the rate capability of silicon nanowires for lithium-ion batteries. *Journal of Power Sources*, 203(0):135 – 139, 2012.

[59] G.R. Goward, N.J. Taylor, D.C.S. Souza, and L.F. Nazar. The true crystal structure of li17m4 (m=ge, sn, pb) - revised from li22m5. *Journal of Alloys and Compounds*, 329(1-2):82 – 91, 2001.

[60] J. Graetz, C. C. Ahn, R. Yazami, and B. Fultz. Nanocrystalline and thin film germanium electrodes with high lithium capacity and high rate capabilities. *Journal of The Electrochemical Society*, 151(5):A698–A702, 2004.

[61] W. A. Grosshans and W. B. Holzapfel. Atomic volumes of rare-earth metals under pressures to 40 gpa and above. *Phys. Rev. B*, 45:5171–5178, Mar 1992.

[62] Bernd Hartke. Global geometry optimization of clusters using genetic algorithms. *The Journal of Physical Chemistry*, 97(39):9973–9976, 1993.

[63] Thomas M. Henderson, Joachim Paier, and Gustavo E. Scuseria. Accurate treatment of solids with the hse screened hybrid. *physica status solidi (b)*, 248(4):767–774, 2011.

[64] R. G. Hennig, A. E. Carlsson, K. F. Kelton, and C. L. Henley. *Ab initio* tizr-ni phase diagram predicts stability of icosahedral tizrni quasicrystals. *Phys. Rev. B*, 71:144103, Apr 2005.

[65] R. G. Hennig, T. J. Lenosky, D. R. Trinkle, S. P. Rudin, and J. W. Wilkins. Classical potential describes martensitic phase transformations between the α, β, and ω titanium phases. *Phys. Rev. B*, 78:054121, Aug 2008.

[66] R. G. Hennig, A. Wadehra, K. P. Driver, W. D. Parker, C. J. Umrigar, and J. W. Wilkins. Phase transformation in si from semiconducting diamond to metallic β-Sn phase in qmc and dft under hydrostatic and anisotropic stress. *Phys. Rev. B*, 82:014101, Jul 2010.

[67] H. Hermann, V. Kokotin, and J. Eckert. Locally fluctuating cooling rate as possible reason for non-crystalline plasticity in metallic glasses. *Europhys. Lett.*, 98(1):16003, 2012.

[68] Jochen Heyd, Juan E. Peralta, Gustavo E. Scuseria, and Richard L. Martin. Energy band gaps and lattice parameters evaluated with the heyd-scuseria-ernzerhof screened hybrid functional. *The Journal of Chemical Physics*, 123(17):174101, 2005.

[69] Jochen Heyd, Gustavo E. Scuseria, and Matthias Ernzerhof. Hybrid functionals based on a screened coulomb potential. *The Journal of Chemical Physics*, 118(18):8207–8215, 2003.

[70] V. Hopf, Herbert Schäfer, and Armin Weiss. Die kristallstruktur der phase Li9 Ge4. *Z. Naturforsch.*, 25b:653, 1970.

[71] M Jansen and J. C. Schon. Rational development of new materials - putting the cart before the horse? *Nature Materials*, 3, 2004.

[72] M Jansen and J Christian Schon. Design in chemican synthesis - an illusion? *Angew. Chem. Int. Ed.*, 45, 2006.

[73] Roy L. Johnston. Evolving better nanoparticles: Genetic algorithms for optimising cluster geometries. *Dalton Trans.*, 2003.

[74] Roy L. Johnston. Evolving better nanoparticles: Genetic algorithms for optimising cluster geometries. *Dalton Trans.*, pages 4193–4207, 2003.

[75] Sung Chul Jung and Young-Kyu Han. Ab initio molecular dynamics simulation of lithiation-induced phase-transition of crystalline silicon. *Electrochimica Acta*, 62(0):73 – 76, 2012.

[76] Kibum Kang, Hyun-Seung Lee, Dong-Wook Han, Gil-Sung Kim, Donghun Lee, Geunhee Lee, Yong-Mook Kang, and Moon-Ho Jo. Maximum li storage in si nanowires for the high capacity three-dimensional li-ion battery. *Applied Physics Letters*, 96(5):053110, 2010.

[77] D. C. Karnopp. Random search techniques for optimization problems. *Automatica*, 1, 1963.

[78] Uday Kasavajjula, Chunsheng Wang, and A. John Appleby. Nano- and bulk-silicon-based insertion anodes for lithium-ion secondary cells. *Journal of Power Sources*, 163(2):1003 – 1039, 2007.

[79] Robert W. Keyes. The electrical properties of black phosphorus. *Phys. Rev.*, 92:580–584, Nov 1953.

[80] Hyunwoo Kim, Chia-Yun Chou, John G. Ekerdt, and Gyeong S. Hwang. Structure and properties of li-si alloys: A first-principles study. *The Journal of Physical Chemistry C*, 115(5):2514–2521, 2011.

[81] Mark D. Kluge, John R. Ray, and Aneesur Rahman. Amorphous-silicon formation by rapid quenching: A molecular-dynamics study. *Phys. Rev. B*, 36:4234–4237, Sep 1987.

[82] Walter Kohn. Nobel lecture: Electronic structure of matter wave functions and density functionals. *Nobel Lectures, Chemistry 1996-2000*, pages 213–237, 1999.

[83] V Kokotin, H Hermann, and J Eckert. Computer simulation of the matrixinclusion interphase in bulk metallic glass based nanocomposites. *J. Phys.: Cond. Matt.*, 23(42):425403, 2011.

[84] G. Kresse and J. Furthmüller. Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set. *Phys. Rev. B*, 54:11169–11186, Oct 1996.

[85] G. Kresse and J. Hafner. *Ab initio* molecular dynamics for liquid metals. *Phys. Rev. B*, 47:558–561, Jan 1993.

[86] G. Kresse and D. Joubert. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B*, 59:1758–1775, Jan 1999.

[87] I. Krivy and B. Gruber. A unified algorithm for determining the reduced (niggli) cell. *Acta Cryst. A*, 32, 1976.

[88] Jing Li and J. R. Dahn. An in situ x-ray diffraction study of the reaction of li with crystalline si. *Journal of The Electrochemical Society*, 154(3):A156–A161, 2007.

[89] Pimpa Limthongkul, Young-Il Jang, Nancy J. Dudney, and Yet-Ming Chiang. Electrochemically-driven solid-state amorphization in lithium-silicon alloys and implications for lithium storage. *Acta Materialia*, 51(4):1103 – 1113, 2003.

[90] Xiao Hua Liu, Shan Huang, S. Tom Picraux, Ju Li, Ting Zhu, and Jian Yu Huang. Reversible nanopore formation in ge nanowires during lithiation-delithiation cycling: An in situ transmission electron microscopy study. *Nano Letters*, 11(9):3991–3997, 2011.

[91] X.J. Liu, Y. Xu, Z.P. Lu, X. Hui, G.L. Chen, G.P. Zheng, and C.T. Liu. Atomic packing symmetry in the metallic liquid and glass states. *Acta Mater.*, 59(16):6480 – 6488, 2011.

[92] L.D. Lloyd and R.L. Johnston. Modelling aluminum clusters with an empirical many-body potential. *Chemical Physics*, 236, 1998.

[93] David C. Lonie and Eva Zurek. Xtalopt: An open-source evolutionary algorithm for crystal structure prediction. *Comput. Phys. Commun.*, 182(2):372 – 387, 2011.

[94] David C. Lonie and Eva Zurek. Identifying duplicate crystal structures: Xtalcomp, an open-source solution. *Comput. Phys. Commun.*, 183(3):690 – 697, 2012.

[95] Andriy O. Lyakhov, Artem R. Oganov, Harold T. Stokes, and Qiang Zhu. New developments in evolutionary structure prediction algorithm {USPEX}. *Comput. Phys. Commun.*, 184(4):1172 – 1182, 2013.

[96] Andriy O. Lyakhov, Artem R. Oganov, and Mario Valle. *Crystal Structure Prediction Using Evolutionary Approach*, pages 147–180. Wiley-VCH, Weinheim, Germany, 2010.

[97] Andriy O. Lyakhov, Artem R. Oganov, and Mario Valle. How to predict very large and complex crystal structures. *Comput. Phys. Commun.*, 181(9):1623 – 1632, 2010.

[98] Scott M. Woodley, Peter D. Battle, Julian D. Gale, and C Richard A. Catlow. The prediction of inorganic crystal structures using a genetic algorithm and energy minimisation. *Phys. Chem. Chem. Phys.*, 1:2535–2542, 1999.

[99] J. Maddox. Crystals from first principles. *Nature*, 335:201, 1988.

[100] Rotem Marom, S. Francis Amalraj, Nicole Leifer, David Jacob, and Doron Aurbach. A review of advanced and practical lithium battery materials. *J. Mater. Chem.*, 21:9938–9954, 2011.

[101] Claire P Massen and Jonathan PK Doye. Power-law distributions for the areas of the basins of attraction on a potential energy landscape. *Physical Review E*, 75(3):037101, 2007.

[102] Andrew J Morris, Chris J Pickard, and R J Needs. Hydrogen/silicon complexes in silicon from computational searches. *Physical Review B*, 78, 2008.

[103] R. J. Nelmes, D. R. Allan, M. I. McMahon, and S. A. Belmonte. Self-hosting incommensurate structure of barium IV. *Phys. Rev. Lett.*, 83(20):4081–4084, 1999.

[104] Reinhard Nesper. Structure and chemical bonding in zintl-phases containing lithium. *Progress in Solid State Chemistry*, 20(1):1 – 45, 1990.

[105] M. N. Obrovac and Leif Christensen. Structural changes in silicon anodes during lithium insertion/extraction. *Electrochemical and Solid-State Letters*, 7(5):A93–A96, 2004.

[106] Artem R. Oganov and Colin W. Glass. Crystal structure prediction using ab initio evolutionary techniques: Principles and applications. *J. Chem. Phys.*, 124, 2006.

[107] Artem R. Oganov and Colin W. Glass. Crystal structure prediction using ab initio evolutionary techniques: Principles and applications. *J.Chem. Phys.*, 124(24):244704, 2006.

[108] Artem R Oganov, Colin W Glass, and Shigeaki Ono. High-pressure phases of $CaCO_3$: Crystal structure prediction and experiment. *Earth and Planetary Science Letters*, 241(1):95–103, 2006.

[109] H. Okamoto. The li-si (lithium-silicon) system. *Journal of Phase Equilibria*, 11:306–312, 1990. 10.1007/BF03029305.

[110] H. Okamoto. Li-si (lithium-silicon). *Journal of Phase Equilibria and Diffusion*, 30(1):118–119, 2009.

[111] Jos F. M. Oudenhoven, Loc. Baggetto, and Peter H. L. Notten. All-solid-

state lithium-ion microbatteries: A review of various three-dimensional concepts. *Advanced Energy Materials*, 1(1):10–33, 2011.

[112] V. Ozolins, C. Wolverton, and Alex Zunger. First-principles theory of vibrational effects on the phase stability of cu-au compounds and alloys. *Physical Review B*, 58, 1998.

[113] J. Paier, M. Marsman, K. Hummer, G. Kresse, I. C. Gerber, and J. G. Ángyán. Screened hybrid density functionals applied to solids. *J. Chem. Phys.*, 124(15):154709, 2006.

[114] Hyoungki Park, Michael R. Fellinger, Thomas J. Lenosky, William W. Tipton, Dallas R. Trinkle, Sven P. Rudin, Christopher Woodward, John W. Wilkins, and Richard G. Hennig. *Ab initio* based empirical potential used to study the mechanical properties of molybdenum. *Phys. Rev. B*, 85:214121, Jun 2012.

[115] Gernot J Pauschenwein. The minimum distance parameterization of crystal lattices. *Journal of Physics A: Mathematical and Theoretical*, 42, 2009.

[116] W. B. Pearson. *Handbook of Lattice Spacing and Structures of Metals,* Vol. II. Pergamon, Oxford, 1967.

[117] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77:3865–3868, Oct 1996.

[118] Chris J. Pickard and R. J. Needs. High-pressure phases of silane. *Phys. Rev. Lett.*, 97, 2006.

[119] Chris J Pickard and R J Needs. Metallization of aluminum hydride at high pressures: A first-principles study. *Physical Review B*, 76, 2007.

[120] Chris J Pickard and R J Needs. Structure of phase iii of solid hydrogen. *Nature Physics*, 3, 2007.

[121] Chris J. Pickard and R. J. Needs. When is h2o not water? *Journal of Chemical Physics*, 127, 2007.

[122] Chris J. Pickard and R. J. Needs. Structures at high pressure from random searching. *Physica Status Solidi*, 246, 2008.

[123] Chris J. Pickard and R. J. Needs. Dense low-coordination phases of lithium. *Physical Review Letters*, 102, 2009.

[124] Chris J. Pickard and R. J. Needs. High-pressure phases of nitrogen. *Physical Review Letters*, 102, 2009.

[125] Chris J Pickard and R J Needs. Stable phases of iron at terapascal pressures. *Journal of Physics: Condensed Matter*, 21(45):452205, 2009.

[126] Chris J Pickard and R J Needs. Ab initio random structure searching. *J. Phys.: Cond. Matt.*, 23(5):053201, 2011.

[127] Peter Pulay. Convergence acceleration of iterative sequences. the case of scf iteration. *Chemical Physics Letters*, 73:393–398, 1980.

[128] V. Raghavan. Al-cu-zr (aluminum-copper-zirconium). *Journal of Phase Equilibria and Diffusion*, 32(5):452–454, 2011.

[129] L. A. Rastrigin. The convergence of the random search method in the extremal control of a many-parameter system. *Authomat. Remote Control*, 24, 1963.

[130] Benjamin C. Revard, William W. Tipton, and Richard G. Hennig. Structure and stability prediction of compounds with evolutionary algorithms. In *Prediction and Calculation of Crystal Structures: Methods and Applications*, Topics in Current Chemistry, pages 1–42. Springer Berlin Heidelberg, 2014.

[131] A.F. Sammells and M.R.S. John. Lithium-germanium electrodes for batteries, August 24 1982. US Patent 4,346,152.

[132] J. Sangster and A.D. Pelton. The ge- li (germanium-lithium) system. *Journal of Phase Equilibria*, 18(3):289–294, 1997.

[133] Johannes Sarnthein, Alfredo Pasquarello, and Roberto Car. Model of vitreous sio_2 generated by an *ab initio* molecular-dynamics quench from the melt. *Phys. Rev. B*, 52:12690–12695, Nov 1995.

[134] Harold A Scheraga. Recent developments in the theory of protein folding: searching for the global energy minimum. *Biophysical Chemistry*, 59, 1996.

[135] J. C. Schon and M. Jansen. Determination of candidate structures for lennard-jones-crystals through cell optimisation. *Ber. Bunsenges. Phys. Chem.*, 98:1541 – 1544, 1994.

[136] J C Schon, M A C Wevers, and M Jansen. Entropically stabilized region on the energy landscape of an ionic solid. *Journal of Physics: Condensed Matter*, 15, 2003.

[137] Wahyu Setyawan and Stefano Curtarolo. High-throughput electronic band structure calculations: Challenges and tools. *Computational Materials Science*, 49(2):299 – 312, 2010.

[138] David A Sholl and Janice A Steckel. *Density Functional Theory: A Practical Introduction*. Wiley, 2009.

[139] G. Simmons and H. Wang. *Single Crystal Elastic Constants and Calculated Aggregate Properties: A Handbook*, 2nd edition. The MIT Press, Cambridge, 1971.

[140] Francisco J Solis and Roger J-B Wets. Minimization by random seearch techniques. *Mathematics of Operations Research*, 6, 1981.

[141] P. Souvatzis, O. Eriksson, M. I. Katsnelson, and S. P. Rudin. Entropy driven stabilization of energetically unstable crystal structures explained from first principles theory. *Physical Review Letters*, 100, 2008.

[142] James C. Spall. *Introduction to Stochastic Search and Optimization*. Wiley, 2003.

[143] Linda A. Stearns, Jan Gryko, Jason Diefenbacher, Ganesh K. Ramachandran, and Paul F. McMillan. Lithium monosilicide (lisi), a low-dimensional silicon-based material prepared by high pressure synthesis: Nmr and vibrational spectroscopy and electrical properties characterization. *Journal of Solid State Chemistry*, 173(1):251 – 258, 2003.

[144] Frank H. Stillinger. Exponential multiplicity of inherent structures. *Physical Review E*, 59, 1999.

[145] Joshua A. Taillon, William W. Tipton, and Richard G. Hennig. Ab initio discovery of novel crystal structure stability in barium under pressure. *In preparation*, 2014.

[146] K Takemura and K Syassen. Pressure-volume relations and polymorphism of europium and ytterbium to 30 gpa. *Journal of Physics F: Metal Physics*, 15(3):543, 1985.

[147] Li Ping Tan, Ziyang Lu, Hui Teng Tan, Jixin Zhu, Xianhong Rui, Qingyu Yan, and Huey Hoon Hng. Germanium nanowires-based carbon composite as anodes for lithium-ion batteries. *Journal of Power Sources*, 206(0):253 – 258, 2012.

[148] William W. Tipton, Clive R. Bealing, Kiran Mathew, and Richard G. Hennig. Structures, phase stabilities, and electrical potentials of li-si battery anode materials. *Phys. Rev. B*, 87:184114, May 2013.

[149] William W. Tipton and Richard G. Hennig. *Random Search Methods*, pages 55–66. Wiley-VCH, Weinheim, Germany, 2010.

[150] William W. Tipton and Richard G. Hennig. GASP - Genetic Algorithm for Structure and Phase Prediction. `http://gasp.mse.cornell.edu`, 2012.

[151] William W Tipton and Richard G Hennig. A grand canonical genetic algorithm for the prediction of multi-component phase diagrams and testing of empirical potentials. *Journal of Physics: Condensed Matter*, 25(49):495401, 2013.

[152] William W. Tipton, Catherine A. Matulis, and Richard G. Hennig. Ab-initio prediction of the li5 ge2 compound. *In preparation*, 2014.

[153] A Togo, F Oba, and I Tanaka. First-principles calculations of the ferroelastic transition between rutile-type and cacl2-type sio2 at high pressures. *Phys. rev. B*, 78:134106, Oct 2008.

[154] Giancarlo Trimarchi, Arthur J. Freeman, and Alex Zunger. Predicting stable stoichiometries of compounds via evolutionary global space-group optimization. *Phys. Rev. B*, 80:092101, Sep 2009.

[155] Giancarlo Trimarchi, Arthur J. Freeman, and Alex Zunger. Predicting stable stoichiometries of compounds via evolutionary global space-group optimization. *Phys. Rev. B*, 80:092101, Sep 2009.

[156] Giancarlo Trimarchi and Alex Zunger. Global space-group optimization

problem: Finding the stablest crystal structure without constraints. *Phys. Rev. B*, 75, 2007.

[157] Mario Valle and Artem R. Oganov. Crystal fingerprint space – a novel paradigm for studying crystal-structure sets. *Acta Cryst. A*, 66(5):507–517, Sep 2010.

[158] Prasana K. Venkatesh, Morrel H. Cohen, Robert W. Carr, and Anthony M. Dean. Bayesian method for global optimization. *Physical Review E*, 55, 1997.

[159] H. G. von Schnering, R. Nesper, K.F. Tebbe, and J. Curda. Structure and properties of $Li_{14}Si_6(Li_{2.33}Si)$, the purple phase, in the system lithium-silicon. *Zeitschrift fur Metallkunde*, 71:357, 1980.

[160] D. Wales. *Energy landscapes with applications to clusters, biomolecules and glasses*. Cambridge University Press, 2003.

[161] David J Wales and Harold A Scheraga. Global optimization of clusters, crystals, and biomolecules. *Science*, 285, 1999.

[162] Wenhui Wan, Qianfan Zhang, Yi Cui, and Enge Wang. First principles study of lithium insertion in bulk silicon. *Journal of Physics: Condensed Matter*, 22(41):415501, 2010.

[163] C. C. Wang and C. H. Wong. Different icosahedra in metallic glasses: Stability and response to shear transformation. *Scripta Mater.*, 66(8):610 – 613, 2012.

[164] C.C. Wang and C.H. Wong. Interpenetrating networks in zrcual and zrcu metallic glasses. *Intermetallics*, 22(0):13 – 16, 2012.

[165] C.C. Wang and C.H. Wong. Structural properties of zrxcu90xal10 metallic glasses investigated by molecular dynamics simulations. *J. Alloys Compd.*, 510(1):107 – 113, 2012.

[166] Junmei Wang, Tingjun Hou, Lirong Chen, and Xiaojie Xu. Automated docking of peptides and proteins by genetic algorithm. *Chenometrics and Intelligent Laboratory Systems*, 45, 1999.

[167] Yanchao Wang, Jian Lv, Li Zhu, and Yanming Ma. Crystal structure prediction via particle-swarm optimization. *Phys. Rev. B*, 82:094116, Sep 2010.

[168] C. John Wen and Robert A. Huggins. Chemical diffusion in intermediate phases in the lithium-silicon system. *Journal of Solid State Chemistry*, 37(3):271 – 278, 1981.

[169] Scott M. Woodley, Peter D. Battle, Julian D. Gale, and C. Richard A. Catlow. The prediction of inorganic crystal structures using a genetic algorithm and energy minimisation. *Phys. Chem. Chem. Phys.*, 1, 1999.

[170] Qiran Xiao, H.W. Sheng, and Yunfeng Shi. Dominant shear bands observed in amorphous zrcual nanowires under simulated compression. *MRS Communications*, 2:13–16, 2 2012.

[171] Ding-Jiang Xue, Sen Xin, Yang Yan, Ke-Cheng Jiang, Ya-Xia Yin, Yu-Guo Guo, and Li-Jun Wan. Improving the electrode performance of ge through ge@c core-shell nanoparticles and graphene networks. *Journal of the American Chemical Society*, 134(5):2512–2515, 2012.

[172] Yansun Yao, John S. Tse, Zhe Song, Dennis D. Klug, Jian Sun, and Yvon Le Page. Structures and superconducting properties of the high-pressure iv and v phases of calcium from first principles. *Phys. Rev. B*, 78:054506, Aug 2008.

[173] Yansun Yao, John S. Tse, and Kaori Tanaka. Metastable high-pressure single-bonded phases of nitrogen predicted via genetic algorithm. *Phys. Rev. B*, 77:052103, Feb 2008.

[174] K. Zhang, P.C. Si, H. Li, Y.F. Li, Y.Y. Jiang, S.L. Zhang, and X.G. Song. Plastic heterogeneity in nanoscale metallic glass. *Physica E*, 44(78):1461 – 1466, 2012.

[175] Xiuwen Zhang, Giancarlo Trimarchi, and Alex Zunger. Possible pitfalls in theoretical determination of ground-state crystal structures: The case of platinum nitride. *Phys. Rev. B*, 79:092102, Mar 2009.

[176] Y. Zhang, N. Mattern, and J. Eckert. Atomic structure and transport properties of cu[sub 50]zr[sub 45]al[sub 5] metallic liquids and glasses: Molecular dynamics simulations. *J. Appl. Phys.*, 110(9):093506, 2011.

[177] Y. Zhang, N. Mattern, and J. Eckert. Study of direct relationship between atomic structures and glass forming abilities of cu[sub 100-x] zr[sub x] (0 ¡= x ¡= 10) liquids by molecular dynamics simulations. *J. Appl. Phys.*, 111(5):053520, 2012.

[178] Y. Zhang, N. Mattern, and J. Eckert. Study of structural anisotropy in cu50zr45al5 metallic glass under uniaxial compression by molecular dynamics simulations. *Intermetallics*, 30(0):154 – 157, 2012.

[179] Y. Zhang, N. Mattern, and J. Eckert. Understanding the relationship between atomic structures and transport properties in (cu0.5zr0.5)100xalx ($\leq$ 10) glass forming liquids: Molecular dynamics simulations. *J. Alloys Compd.*, 514(0):141 – 149, 2012.

[180] Kejie Zhao, Wei L. Wang, John Gregoire, Matt Pharr, Zhigang Suo, Joost J. Vlassak, and Efthimios Kaxiras. Lithium-assisted plastic deformation of silicon electrodes in lithium-ion batteries: A first-principles theoretical study. *Nano Letters*, 11(7):2962–2967, 2011.

[181] Anatoly Zhigljavsky and Antanas Zilinskas. *Stochastic Global Optimization*. Springer, 2007.

[182] Houlong L. Zhuang, Arunima K. Singh, and Richard G. Hennig. Computational discovery of single-layer iii-v materials. *Phys. Rev. B*, 87:165415, Apr 2013.